

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 15:08:29 ; Search time 46 Seconds

(without alignments)  
1535.507 Million cell updates/sec

Title: US-10-339-656-2

Perfect score: 2410

Sequence: 1 MCFLEASHSVEDTHPSHY.....QRTLEALHPFDLKKK 445

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq.19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2410	100.0	445	23	AAE20170	Human kinase prote
2	2294.5	95.2	481	22	AAE69705	Human cell cycle r
3	2177.5	90.4	479	19	AAW49914	Mouse CLK serine/c
4	2127.5	88.3	499	22	AAE65648	Novel protein kina
5	2117.5	87.9	451	20	AAV27054	Human protein kina
6	1887	78.3	484	23	AAU75108	Serine and Threoni
7	1808.5	75.0	482	19	AAW49911	Mouse CLK serine/c
8	1535	63.7	283	21	AAE33778	Human secreted pro
9	1526	63.3	350	21	AAE58831	Breast and ovarian

10	1517	62.9	283	21	AAE33777	Human secreted pro
11	1400.5	58.1	352	23	AAE20171	Human kinase prote
12	1331.5	55.2	497	19	AAW49912	Mouse CLK serine/c
13	1314.5	54.5	490	22	AAE23354	Novel human diapo
14	1310.5	54.4	431	22	AAE65506	Human protein kina
15	1223.5	50.8	485	19	AAW49913	Mouse CLK serine/c
16	1222	50.7	511	22	AAE58832	Drosophila melanog
17	1222	50.7	511	22	AAE65924	Drosophila melanog
18	1051	43.6	265	24	ABR41275	Human DITP intrac
19	848.5	35.2	425	21	AAE29704	Arabidopsis thalia
20	848.5	35.2	427	21	AAE29703	Arabidopsis thalia
21	848.5	35.2	435	21	AAE29702	Arabidopsis thalia
22	811.5	33.7	398	21	AAE41256	Arabidopsis thalia
23	811.5	33.7	464	21	AAE41255	Arabidopsis thalia
24	811.5	33.7	478	21	AAE41254	Arabidopsis thalia
25	748	31.0	335	21	AAE51301	Arabidopsis thalia
26	748	31.0	345	21	AAE51300	Arabidopsis thalia
27	748	31.0	348	21	AAE51299	Arabidopsis thalia
28	550.5	22.8	527	23	ABE80743	Rat YAK1 protein k
29	550.5	22.8	551	22	AAE02843	Mouse serine/threo
30	549.5	22.8	526	23	ABE30696	Murine mdmR2 poly
31	549.5	22.8	549	21	AAE68781	Amino acid sequenc
32	546.5	22.7	528	19	AAE64559	Human protein kina
33	546.5	22.7	528	22	AAE02011	Human YAK1 (YAK1)
34	544.5	22.6	568	19	AAW75791	Human protein kina
35	544.5	22.6	568	21	AAE00059	Human protein kina
36	544.5	22.6	568	22	AAE02013	Human YAK3b (YAK3
37	544.5	22.6	588	19	AAW75790	Human protein kina
38	544.5	22.6	588	21	AAE00058	Human YAK3-2 alpha poly
39	544.5	22.6	588	21	AAE68792	Human YAK3 (YAK3)
40	544.5	22.6	588	22	AAE02012	Human YAK3 (YAK3)
41	536.5	22.3	553	22	AAE65662	Novel protein kina
42	532.5	22.1	722	22	ABE60747	Drosophila melanog
43	517	21.5	620	21	AAE58052	Human protein kina
44	505	21.0	539	22	AAU04204	Drosophila melanog
45	501.5	20.8	527	22	ABE63499	Drosophila melanog

#### ALIGNMENTS

RESULT 1	AAE20170	standard; Protein; 445 AA.
ID	AAE20170	
XX	AAE20170:	
AC	18-JUN-2002 (first entry)	
XX		
DT		
XX		
DE	Human kinase protein.	
XX		
KW	Human; kinase protein; enzyme; cytosolic; osteopathic; gene expression;	
KW	colon-moderately differentiated adenocarcinoma; chromosome mapping; gene;	
KW	drug screening; therapeutic; gene therapy; tissue typing; chromosome 5;	
KW	bone osteosarcoma.	
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
FT	Modified-site	10..13 "Casein kinase II phosphorylation site"
FT	Modified-site	24..31 "Tyrosine kinase phosphorylation site"
FT	Modified-site	25..28 "Tyrosine kinase phosphorylation site"
FT	Modified-site	29..36 "Casein kinase II phosphorylation site"
FT	Modified-site	55..61 "Tyrosine kinase phosphorylation site"
FT	Modified-site	69..71 "Tyrosine kinase phosphorylation site"
FT	Modified-site	72..74 "Protein kinase C phosphorylation site"
FT	Modified-site	/note="Protein kinase C phosphorylation site"

FT Modified-site 73..76 /note= "cAMP and cGMP-dependent protein kinase phosphorylation site"  
 FT Modified-site 76..78 /note= "Protein kinase C phosphorylation site"  
 FT Modified-site 94..96 /note= "Protein kinase C phosphorylation site"  
 FT Modified-site 97..100 /note= "Protein kinase C phosphorylation site"  
 FT Modified-site 102..105 /note= "cAMP and cGMP-dependent protein kinase phosphorylation site"  
 FT Modified-site 128..131 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 129..153 /note= "Casein kinase II phosphorylation site"  
 FT Binding-site 147..152 /note= "Protein kinase ATP-binding region signature"  
 FT Modified-site 176..179 /note= "N-myristoylation site"  
 FT Modified-site 209..212 /note= "Asn is N-glycosylated"  
 FT Modified-site 246..258 /note= "Casein kinase II phosphorylation site"  
 FT Active-site 247..250 /note= "Serine/threonine protein kinase active-site signature"  
 FT Modified-site 277..279 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 292..295 /note= "Protein kinase C phosphorylation site"  
 FT Modified-site 303..305 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site 324..344 /note= "Protein kinase C phosphorylation site"  
 FT Region /note= "Helix 1"  
 FT Modified-site 368..370 /note= "Protein kinase C phosphorylation site"  
 FT Modified-site 425..427 /note= "Protein kinase C phosphorylation site"  
 FT Modified-site 429..432 /note= "Casein kinase II phosphorylation site"  
 FT Modified-site /note= "Casein kinase II phosphorylation site"  
 XX MO200216567-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 24-AUG-2001; 2001WO-US26389.  
 XX  
 PR 24-AUG-2000; 2000US-227470P.  
 PR 19-MAR-2001; 2001US-0810671.  
 XX  
 PA (APPL-) APPLERA CORP.  
 PI Yan C, Ye J, Ketchum KA, Di Francesco V, Beasley EW,  
 PI WPI; 2002-269354/31.  
 DR N-Psdb; AAD32038.  
 XX  
 PT New human kinase proteins and nucleic acids, useful in drug screening  
 PT assays, identifying modulators of kinase activity or treating disorders  
 PT characterized by absence or unwanted expression of the protein  
 XX  
 PS Claim 1; Fig 2; 81pp; English.  
 XX  
 CC The invention relates to isolated human kinase proteins and nucleic  
 CC acids. The nucleic acid and peptide sequences can be used as models for  
 CC the development of human therapeutic targets, aid in the identification  
 CC of therapeutic proteins and serve as targets for the development of  
 CC human therapeutic agents that modulate kinase activity in cells and  
 CC tissues that express the kinase. The nucleic acids are useful as probes  
 CC or primers, in constructing recombinant vectors, for expressing  
 CC antigenic portions of the proteins, chromosome mapping, drug screening,  
 CC testing an individual for a genotype, and for gene therapy in patients

CC containing cells that are aberrant in kinase gene expression. The  
 CC proteins may be used in drug screening assays, in the identification of  
 CC compounds that modulate, stimulate or inhibit kinase activity, in  
 CC pharmacogenomic analysis, in treating disorders characterized by an  
 CC absence or unwanted expression of the protein (bone osteosarcoma, or  
 CC colon/moderately differentiated adenocarcinoma), and in generating  
 CC antibodies specific for the peptides. Such antibodies can be used to  
 CC detect the protein in situ, in vitro, or in cell lysate or supernatant,  
 CC to isolate and purify the proteins from host cells, pharmacogenomic  
 CC analysis, tissue typing, and in inhibiting protein function. The present  
 CC sequence is human kinase protein. Human kinase protein gene is  
 CC located on chromosome 5.  
 XX  
 SQ Sequence 445 AA;  
 QY  
 Db 1 MCIPLEASHSVEDTTPSHYLEARSINERDYPDRRYVDEYRNDCYCEGVPRHYHRIESG 60  
 1 MCIPLEASHSVEDTTPSHYLEARSINERDYPDRRYVDEYRNDCYCEGVPRHYHRIESG 60  
 QY 61 YRIHCKSSVSRSSPKRRNRHSCSHQSRSKSHRRKRSRS1EDDEGHILICSGDVLK 120  
 61 YRIHCKSSVSRSSPKRRNRHSCSHQSRSKSHRRKRSRS1EDDEGHILICSGDVLK 120  
 Db 121 ARREIVDTLGEAGFGVVCIDHGMGMVAIVVNGVRYEAASEIQLVLEHNSNDP 180  
 121 ARREIVDTLGEAGFGVVCIDHGMGMVAIVVNGVRYEAASEIQLVLEHNSNDP 180  
 QY 122 ARREIVDTLGEAGFGVVCIDHGMGMVAIVVNGVRYEAASEIQLVLEHNSNDP 180  
 122 ARREIVDTLGEAGFGVVCIDHGMGMVAIVVNGVRYEAASEIQLVLEHNSNDP 180  
 Db 121 ARREIVDTLGEAGFGVVCIDHGMGMVAIVVNGVRYEAASEIQLVLEHNSNDP 180  
 121 ARREIVDTLGEAGFGVVCIDHGMGMVAIVVNGVRYEAASEIQLVLEHNSNDP 180  
 QY 181 NSVFCVQMLEWFDHGHVCIYFELGLSTYDIKENSFLPQIDHIRQMAVQICOSINF 240  
 181 NSVFCVQMLEWFDHGHVCIYFELGLSTYDIKENSFLPQIDHIRQMAVQICOSINF 240  
 Db 181 NSVFCVQMLEWFDHGHVCIYFELGLSTYDIKENSFLPQIDHIRQMAVQICOSINF 240  
 181 NSVFCVQMLEWFDHGHVCIYFELGLSTYDIKENSFLPQIDHIRQMAVQICOSINF 240  
 QY 241 LHHNKLTHTDLPENILFYKSDYVVKYNSKMRDEPTLANTDIKVDPGSAIYDDEHST 300  
 241 LHHNKLTHTDLPENILFYKSDYVVKYNSKMRDEPTLANTDIKVDPGSAIYDDEHST 300  
 Db 241 LHHNKLTHTDLPENILFYKSDYVVKYNSKMRDEPTLANTDIKVDPGSAIYDDEHST 300  
 241 LHHNKLTHTDLPENILFYKSDYVVKYNSKMRDEPTLANTDIKVDPGSAIYDDEHST 300  
 QY 301 LVSTRHYRAPEVILAGMSQPCDWSIGCILLEYLGFVFOYHDSKEHLLAMMERILGPI 360  
 301 LVSTRHYRAPEVILAGMSQPCDWSIGCILLEYLGFVFOYHDSKEHLLAMMERILGPI 360  
 Db 301 LVSTRHYRAPEVILAGMSQPCDWSIGCILLEYLGFVFOYHDSKEHLLAMMERILGPI 360  
 301 LVSTRHYRAPEVILAGMSQPCDWSIGCILLEYLGFVFOYHDSKEHLLAMMERILGPI 360  
 QY 361 POHMIQTKRKRYFHHNQDMDDEHSAGRYVRRCKPLKEFMLCHDEBEHKLFDIVRRML 420  
 361 POHMIQTKRKRYFHHNQDMDDEHSAGRYVRRCKPLKEFMLCHDEBEHKLFDIVRRML 420  
 Db 361 POHMIQTKRKRYFHHNQDMDDEHSAGRYVRRCKPLKEFMLCHDEBEHKLFDIVRRML 420  
 361 POHMIQTKRKRYFHHNQDMDDEHSAGRYVRRCKPLKEFMLCHDEBEHKLFDIVRRML 420  
 QY 421 EYDPTQRTITLDEALQHPFDLLKKK 445  
 421 EYDPTQRTITLDEALQHPFDLLKKK 445  
 Db 421 EYDPTQRTITLDEALQHPFDLLKKK 445  
 421 EYDPTQRTITLDEALQHPFDLLKKK 445  
 RESULT 2  
 ID AAB69705 standard; Protein; 481 AA.  
 XX  
 AC AAB69705;  
 XX  
 AC 14-AUG-2001 (first entry)  
 XX  
 DE Human cell cycle regulating protein 53.  
 XX  
 KW Human; cell cycle regulating protein 53; cancer; blood disease; HIV;  
 KW immunological disease; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200130833-A1.  
 XX  
 PD 03-MAY-2001.  
 XX  
 PF 16-OCT-2000; 2000WO-CN00328.  
 XX



```

PR 22-OCT-1999; 99CN-0119816.
XX
XX (SHAN-) SHANGHAI BIO ROAD GENE DEV LTD.
XX
XX Mao Y, Xie Y;
XX
XX WPI; 2001-300480/31.
XX
XX N-PSDB; AAF99402.
XX
XX New cell cycle-regulating protein 53 and its polynucleotide, applicable
XX in diagnosis and treatment of malignant tumour, haemopathy, human
XX immunodeficiency virus infection, immunological diseases and various
XX inflammation
XX
XX Claim 1; Page 23-24; 30pp; Chinese.
XX
XX The present invention provides the protein and coding sequences for human
XX cell cycle regulating protein 53. The sequences can be used in the
XX diagnosis and treatment of malignant tumours, haemopathy, human
XX immunodeficiency virus (HIV) infection, immunological diseases and
XX various types of inflammation. The present sequence is the cell cycle
XX regulating protein 53.
XX
XX Sequence 481 AA;
SQ
Query Match 95.2%; Score 2294.5; DB 22; Length 481;
Best Local Similarity 95.5%; Pred. No. 2.3e-210;
Matches 429; Conservative 2; Mismatches 7; Indels 11; Gaps 2;
QY 8 SH-SVEEDTH-----PSHYLEARSLENERDYRDRRYVDEYRNDYCEGYVPRHYRD 56
DB 33 SHSSTQENRHKCPHHQFKESDCHYLEARSLNERDYRDRRYVDEYRNDYCEGYVPRHYRNG 92
QY 57 IESGRIHCSKSSVRSRSPKRNKRNHCSSHQSSKSHRRKRSISIEDDEGHILCOQG 116
DB 93 IESGYRIHCSKSSVRSRSPKRNKRNHCSSHQSSKSHRRKRSISIEDDEGHILCOQG 152
QY 117 DVLRARYEIVDTLGGAGFGKVEECIDHGMDGQHVAVKIVKNGYRFAARSEIOVLEHNL 176
DB 153 DVLRARYEIVDTLGGAGFGKVEECIDHGMDGQHVAVKIVKNGYRFAARSEIOVLEHNL 212
QY 177 STDPSVFRVCVOMLEMFHDHGVCIFFELGLSTVDYFIKENSFLFPQIDHROMAYOICO 236
DB 213 STDPSVFRVCVOMLEMFHDHGVCIFFELGLSTVDYFIKENSFLFPQIDHROMAYOICO 272
QY 227 SINFLHNKLTHTTDLKPNILFVKSQDYVYKYNKMKRDERTLKNTDIKAVDFGSAATYDDE 296
DB 273 SINFLHNKLTHTTDLKPNILFVKSQDYVYKYNKMKRDERTLKNTDIKAVDFGSAATYDDE 332
QY 297 HHSITVSTRHRAPEVILALGMSOPCDWISIGCIIIEVYLGFTVFQTHDSKEHLAMMERI 356
DB 333 HHSITVSTRHRAPEVILALGMSOPCDWISIGCIIIEVYLGFTVFQTHDSKEHLAMMERI 392
QY 357 LGPIPOHMIQKTRKRRKYPFHNNQIDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLFYDIV 416
DB 393 LGPIPOHMIQKTRKRRKYPFHNNQIDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLFYDIV 452
QY 417 RRMLEYDPTORTITLDEALQHPFDLLKKX 445
DB 453 RRMLEYDPTORTITLDEALQHPFDLLKKX 481

```

```

KW signal transduction; cancer; contraceptive; mouse; therapy;
XX diagnosis.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
XX Domain 29..44 /note="nuclear localisation domain"
XX Domain 159..476 /note="catalytic domain"
XX Peptide 384..389 /note="LAMMER motif"
XX
XX WO97/48723-A2.
XX
XX 24-DEC-1997.
XX
XX 17-JUN-1997; 97WO-IB00946.
XX
XX 19-DEC-1996; 96US-0034286.
XX 17-JUN-1996; 96US-0019629.
XX 09-AUG-1996; 96US-0023485.
XX 13-NOV-1996; 96US-0030860.
XX 15-NOV-1996; 96US-0030964.
XX
XX (PLAC ) MAX PLANCK GRS FOERDERUNG WISSENSCHAFTEN.
XX Aoki N, Chen Z, Kharitonov AI, Kim YW, Nayler O;
XX Ullrich A, Wang HY;
XX WPI; 1998-120302/11.
XX
XX New phosphatase and kinase enzyme(s) - useful in the diagnosis and
XX treatment of signal transduction disorders
XX
XX Claim 11; Fig 4; 138pp; English.
XX
XX This polypeptide comprises novel mouse CLK serine/threonine kinase
XX mCLK4, from the CLK serine/threonine kinase family of proteins that
XX regulate RNA splicing in cells. mCLK4 cDNA was cloned from a mouse
XX embryo 11.5 p.c. 12AP cDNA library. The invention discloses the
XX discovery of novel protein kinases mCLK2 (see AAM4912), mCLK3 and
XX mCLK4 (see AAM4914) of mol.wt. 59.9, 58.5 and 57.2 kDa,
XX respectively, as well as other novel proteins (see AAM4906-10)
XX involved in cellular signal transduction, and provides vectors,
XX host cells, purified recombinant proteins, methods for identifying
XX compounds that activate or inhibit the novel proteins, as well as
XX methods for the diagnosis and treatment of diseases associated with
XX the novel proteins. Overexpression of CLK serine/threonine kinases
XX has been implicated in certain types of cancer. Compounds that
XX inhibit their catalytic activity or disrupt their interactions
XX with natural binding partners may act as anti-cancer therapeutics.
XX mCLK related molecules and compounds may also be useful as male
XX contraceptives.
XX
XX Sequence 479 AA;
SQ
Query Match 90.4%; Score 2177.5; DB 19; Length 479;
Best Local Similarity 91.1%; Pred. No. 3.4e-199;
Matches 409; Conservative 9; Mismatches 18; Indels 13; Gaps 4;
QY 8 SH-SVEEDTH-----PSHYLEARSLENERDYRDRRYVDEYRNDYCEGYVPRHYRD 56
DB 33 SHSSTQENRHKCPHHQFKESDCHYLEARSLNERDYRDRRYVDEYRNDYCEGYVPRHYRD 91
QY 57 IESGRIHCSKSSVRSRSPKRNKRNHCSSHQSSKSHRRKRSISIEDDEGHILCOQG 116
DB 92 VESTYRIHCSKSSVRSRSPKRNKRNHCSSHQSSKSHRRKRSISIEDDEGHILCOQG 151
QY 117 DVLRARYEIVDTLGGAGFGKVEECIDHGMDGQHVAVKIVKNGYRFAARSEIOVLEHNL 176
DB 152 DVLRARYEIVDTLGGAGFGKVEECIDHGMDGQHVAVKIVKNGYRFAARSEIOVLEHNL 211

```

```

OY 177 STDNSVFRVCVQMLEMPDHGHVCIIVFELLGLSTYDPIKENSFLPFQIDHIRQAVYQICQ 236
DB 212 STDPSVFRVCVQMLEMPDHGHVCIIVFELLGLSTYDPIKENSFLPFQIDHIRQAVYQICQ 271
OY 237 SINFLHNKLTHTPLKPNILFVKSDDYVVKNSKKRBERLTAKNDIVVPGSATYDDE 296
DB 212 SINFLHNKLTHTPLKPNILFVKSDDYVVKNSKKRBERLTAKNDIVVPGSATYDDE 330
OY 297 HNSTLSTRHRAPEVILALGMSQPCDVMSIGCIIIEYVLGFTVPGTDSKEHLAMERI 356
DB 331 HNSTLSTRHRAPEVILALGMSQPCDVMSIGCIIIEYVLGFTVPGTDSKEHLAMERI 390
OY 357 LGPIPOHMIQKTRKRYFHNQDMDDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 416
DB 391 LGPIPAHMIQKTRKRYFHNQDMDDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 450
OY 417 RRMLEYDPTORITLDEALQHPFDLLKKK 445
DB 451 RRMLEYDPTORITLDEALQHPFDLLKKK 479

RESULT 4
AAB65648
ID AAB65648 standard; Protein: 499 AA.
AC AAB65648;
XX
XX
XX 27-MAR-2001 (first entry)
DT
XX
XX Novel protein kinase, SEQ ID NO: 175.
DE
XX
XX Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
XX immunosuppressive; carbant; renal; antiinflammatory; antiaesthetic;
XX dermatological; antidiabetic; antifertility; gene therapy; vaccine;
XX immune disorder; cardiovascular disease; neurodegenerative disease;
XX cancer; autoimmune disorder; stroke; inflammatory bowel disease;
XX inflammatory pelvic disease; multiple sclerosis; psoriasis.
XX Homo sapiens.
OS
XX MO200073469-A2.
FN
XX
XX 07-DEC-2000.
PD
XX
XX 26-MAY-2000; 2000QWO-US14842.
PF
XX
XX 28-MAY-1999; 99US-0136503.
PR
XX
XX (SUGC-) SUGEN INC.
PA
XX
XX Plowman GD, Martinez R, Whyte D, Sudersanam S;
PI
XX WPI; 2001-032161/04.
DR
XX N-PSDB; AAF44675.
XX
XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
XX treating immune-related diseases and disorders, cardiovascular disease,
XX neurodegenerative diseases and/or cancers -
XX
XX Claim 10; Fig 1; 310pp; English.
XX
XX The present sequence is a novel protein kinase. The novel protein kinases
XX and the nucleic acids that encode them may be used in the treatment and
XX diagnosis of diseases associated with inappropriate kinase expression
XX such as immune-related diseases and disorders, cardiovascular disease,
XX neurodegenerative diseases and/or cancers. The nucleic acids and
XX complementary sequences may also be used as DNA probes in diagnostic
XX assays. The kinase polypeptides may be used as antigens in the production
XX of antibodies of kinase expression and activity. Anti-kinase antibodies
XX and kinase antagonists may also be used to down regulate kinase
XX expression and activity. Diseases related to kinase expression and
XX activity include rheumatoid arthritis, atherosclerosis, autoimmune
XX disorders, complications of organ transplantation, myocardial infarction,

```

```

CC immune disorders, cardiomyopathies, strokes, renal failure,
CC oxidative-stress related disorders, chronic inflammatory bowel disease,
CC chronic inflammatory pelvic disease, multiple sclerosis, asthma,
CC osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
CC reproductive disorders.
XX
XX Sequence 499 AA;
SQ
Query Match 88.3%; Score 2127.5; DB 22; Length 499;
Best Local Similarity 89.3%; Pred. No. 2.2e-194;
Matches 401; Conservative 2; Mismatches 5; Indels 41; Gaps 3;
OY 8 SH-SVEBDTH-----PSHYLEARSLNERDYRDRYVDEYNDYCEGYVPRHYRD 56
DB 81 SHSSTQENRHCPRHNFQKESDCHYLEARSLNERDYRDRYVDEYNDYCEGYVPRHYRD 140
OY 57 IESGYRIHCSKSSVRSRRSSPKRRKRNHSCSHQSRSSHKRRKRSIJDDEBGLICQSG 116
DB 141 IESGYRIHCSKSSVRSRRSSPKRRKRNHSCSHQSRSSHKRRKRSIJDDEBGLICQSG 176
OY 117 DVLRARYEIVDTLGEAGFAGKVVCEIDHGMGMHVAVKIVKRVGRYRARSSEIOVLEHLN 176
DB 177 -----XEVDTLGEAGFAGKVVCEIDHGMGMHVAVKIVKRVGRYRARSSEIOVLEHLN 230
OY 177 STDNSVFRVCVQMLEMPDHGHVCIIVFELLGLSTYDPIKENSFLPFQIDHIRQAVYQICQ 236
DB 231 STDNSVFRVCVQMLEMPDHGHVCIIVFELLGLSTYDPIKENSFLPFQIDHIRQAVYQICQ 290
OY 237 SINFLHNKLTHTPLKPNILFVKSDDYVVKNSKKRBERLTAKNDIVVPGSATYDDE 296
DB 291 SINFLHNKLTHTPLKPNILFVKSDDYVVKNSKKRBERLTAKNDIVVPGSATYDDE 350
OY 297 HNSTLSTRHRAPEVILALGMSQPCDVMSIGCIIIEYVLGFTVPGTDSKEHLAMERI 356
DB 351 HNSTLSTRHRAPEVILALGMSQPCDVMSIGCIIIEYVLGFTVPGTDSKEHLAMERI 410
OY 357 LGPIPOHMIQKTRKRYFHNQDMDDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 416
DB 411 LGPIPOHMIQKTRKRYFHNQDMDDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 470
OY 417 RRMLEYDPTORITLDEALQHPFDLLKKK 445
DB 471 RRMLEYDPTORITLDEALQHPFDLLKKK 499

RESULT 5
AA27054
ID AA27054 standard; Protein: 451 AA.
XX
XX AA27054;
AC
XX
XX 08-OCT-1999 (first entry)
DT
XX
XX Human protein kinase (HPKM)-3 (clone ID 339963).
DE
XX
XX Human protein kinase molecule; HPKM; human; protein kinase;
XX phosphate group; cancer; immune disorder.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX Modified-site 182
XX FT /note= "potential N-glycosylation site"
XX Domain 135..159
XX FT /note= "potential signature sequence for protein
XX FT kinase catalytic domain"
XX Domain 252..264
XX FT /note= "potential signature sequence for protein
XX FT kinase catalytic domain"
XX Modified-site 4
XX FT /note= "potential phosphorylation site"
XX FT Modified-site 22
XX FT /note= "potential phosphorylation site"

```

Query Match	87.9%;	Score 217.5;	DB 20;	Length 451;
Best Local Similarity	88.9%;	Pred. No. 1.7e-193;		
Matches 399;	Conservative 3;	Mismatches 6;	Indels 41;	Gaps 3

OY		8	SH-SVEEDT-----PSHYLEKASINENDDYDRAYVDERYNDYCEGVPRRYHND	56
Dd		33	SHSTQENRCKCPHHQKESDCCHYLEARRSLNENDYDRIYVDYRDDYCEGIVPRRYHD	92
OY		57	IETGYRTICKSSSVRSRRSPKRKRNRHCSSHOSRSKSHRRKRSIIEDEEGHLICQSG	111
Dd		93	IETGYRTICKSSSVRSRRSPKKRKRRHCSSHOSR-----	122
OY		117	DVLARARELYDTLGECAFNGVVECIDHGMDGMHVAVKIYNNGRYREAAARSEIQULEHN	177
Dd		129	-----WKSDVTDLGEAGFGKVCEIDHGMDGMHVAIKYNVGRYREAAARSEIQULEHN	187
OY		177	STDENSVFRCOMLEWFDHGHVCIVFELLGLSTYDFIKENSFLPPOIHIROMAYOICO	233
Dd		183	STDNSNVFRCYMLEWFDHGHVCIVFELLGLSTYDFIKENSFLPPOIHIRMAYOICO	240
OY		237	SINFLENKLTHFDLPKENILPVKSIDYVVKYKSKMRDERTLKNYDIKYVDFGSATYDDE	293
Dd		243	SINFLENKLTHFDLPKENILPFYSIDYVVKYKSKMRDERTLKNYDIKYVDFGSATYDDE	300
OY		297	HHSFLVSTRHARAPEVLLAIGMOPDWISIGILLLEYLGFTVPTHDSKEHLAMERI	353
Dd		303	HHSFLVSTRHARAPEVLLAIGMSQPCDWSIGILLLEYLGFTVPTHDSKEHLAMERI	360
OY		357	LGPFPQMNIOTKTRKKYEHHNQLDMDHESSAGHYVRROCKPLKEFMLCHDEHEKLFDIV	411
Dd		363	LGPFPQMNIOTKTRKKYFHFNQLDMDHESSAGHYVRROCKPLKEFMLCHDEHEKLFDIV	422
OY		417	RRLMEYDPFORITLDEALOHPPFDLLKKK	445
Dd		423	RRLMEYDPFORITLDEALOHPPFDLLKKK	451
<hr/>				
RESULT 6				
AAU75108				
ID	AAU75108	standard; Protein;	484 AA.	
XX				
AC	AAU75108;			
XX				
DT	09-APR-2002	(first entry)		
DE				
XX		Serine and Threonine protein kinase CLK.		
KW	MARPAP-K3; AP-3 delta; App-695; amyloid A4; Hsp9; heat shock protein 8;			
KW	L130; NY-REN-58; P38 Alpha; P38 Beta; ERK3; KIAA0934; CDK3;			
KW	cell cycle dependent kinase 9; CLK; C-NAP-1; clathrin heavy chain;			
KW	amphiphysin; PNS109; KIAA106; neurodegenerative disorder;			
KW	non-insulin dependent diabetes mellitus; NIDDM; Alzheimer's disease;			
KW	inflammatory disease; rheumatoid arthritis; inflammatory bowel disorder;			
KW	atherosclerosis; cardiac hypertrophy; hypoxic brain injury;			
KW	yeast two-hybrid; signal transduction pathway; human;			
KW	mitogen activated protein kinase.			
OS	Homo sapiens.			
XX				
XX				
EH	Key	Location/Qualifiers		
FT	Region	1..364		
FT	/note=	"This region binds the mitogen activated protein		
FT		kinase (MAPK) ERK3 (see AAU75105), the bait		
FT		protein in a yeast two-hybrid assay, producing		
FT		a complex claimed in claim 1"		
XX				
PN	WC0200198524-A2.			
XX				
PD	27-DEC-2001.			
PF	21-JUN-2001; 2001WO-US19762.			
XX				
PR	22-JUN-2000; 2000US-213245P.			
PA	(MYRI-) MYRIAD GENETICS INC.			
XX				



```
Db      48 HSKTTD---SYLSESRINEKAYHSRRYVDEYRNDYM-GYEPGHPYGEPCGRYOMSSKS 103
Qy      69 SVASRRSPKPK-RNR-HCSSHOSRSKSHRRKRSRS-IEDDEBGLICQSGDVLARYEIV 126
Db      104 SGRSGRSSYKSKSRSHHTSOHSHSDGSHSRKRSRSVDEDEBGLICQSGDVLARSYEIV 163
Qy      127 DTGEGAFGKVCIECIDHGMGMHAYVIVKVGRRYRARSSEIOVLEHNLSTDPNSVFC 186
Db      164 DTGEGAFGKVCIECIDHGMGMHAYVIVKVGRRYRARSSEIOVLEHNLSTDPNSVFC 223
Qy      187 VQMLEWFDHGHVCIYFELLGLSTYDFIKENSFLPFQIDIRQMAVOICOSINFLHNKL 246
Db      224 VQMLEWFEHRGHGHTCIYFELLGLSTYDFIKENSFLPFQIDIRQMAVOICOSINFLHNKL 283
Qy      247 THTDLPKPNILFYKSDIVYVKNSKMRDERTLNKNTDIKVVDPGSAFYDDEHNSTLVSTRH 306
Db      284 THTDLPKPNILFYKSDIVYVKNSKMRDERTLNKNTDIKVVDPGSAFYDDEHNSTLVSTRH 342
Qy      307 YRAPEVILALGWSQPCDWSIGCILLIYYLGFVFPCHDKEHLAMMERILGLPIQHMIO 366
Db      343 YRAPEVILALGWSQPCDWSIGCILLIYYLGFVFPCHDKEHLAMMERILGLPIQHMIO 402
Qy      367 KTRKRYFPHNOLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLFDLVRMLBYDPFQ 426
Db      403 KTRKRYFPHNOLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLFDLVRMLBYDPFQ 462
Qy      427 RITLDEALQHPFDLLKK 444
Db      463 RITLKEALKHPFPYPLKK 480

RESULT 8
AAB33778
ID AAB33778 standard; Protein; 283 AA.
XX
XX AAB33778;
XX
XX 02-FEB-2001 (first entry)
XX
DE Human secreted protein BLAST search protein SEQ ID NO: 122.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
XX vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX
XX Homo sapiens.
XX
XX WO200056753-A1.
XX
XX 28-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-US06765.
XX
XX 23-MAR-1999; 99US-0126051.
XX
XX 10-DEC-1999; 99US-0169906.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI, 2000-594570/56.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
XX preventing, treating or ameliorating a disorder -
XX
XX Disclosure, Page 400-401; 410pp; English.
XX
XX The invention relates to the isolation of genes AAC59277-C59325 encoding
XX the human secreted proteins AAB33718-B33764. The sequence is used as a
XX query sequence for doing BLASTX searches to identify homologous
XX sequences. The genes and proteins are useful for preventing,
```

```
CC ameliorating or treating medical conditions, e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal
CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
CC disease, allergies, autoimmune haemolytic anaemia, autoimmune
CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular
CC disorders such as myocardial ischaemias; (d) wound healing; (e)
CC neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections.
XX
XX Sequence 283 AA;
XX
XX
XX Query Match 63.7%; Score 1535; DB 21; Length 283;
XX Best Local Similarity 100.0%; Pred. No. 4, 2e-138;
XX Matches 283; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 163 EAARSEIOVLEHNLSTDPNSVFCVQMLEWFDHGHVCIYFELLGLSTYDFIKENSFLPF 222
XX 1 EAARSEIOVLEHNLSTDPNSVFCVQMLEWFDHGHVCIYFELLGLSTYDFIKENSFLPF 60
XX
XX 223 QIDHIRQMAVOICOSINFLHNKL/THTDLPKPNILFYKSDIVYVKNSKMRDERTLNKNTD 282
XX 61 QIDHIRQMAVOICOSINFLHNKL/THTDLPKPNILFYKSDIVYVKNSKMRDERTLNKNTD 120
XX
XX 283 IKVVDPGSATYDDEHNSTLVSTRHRAPEVILALGWSQPCDWSIGCILLIYYLGFVFPQ 342
XX 121 IKVVDPGSATYDDEHNSTLVSTRHRAPEVILALGWSQPCDWSIGCILLIYYLGFVFPQ 180
XX
XX 343 THDSKEHLAMMERILGLPIQHMIOKTRKRYFPHNOLDMDHSSAGRYVRRCKPLKEFM 402
XX 181 THDSKEHLAMMERILGLPIQHMIOKTRKRYFPHNOLDMDHSSAGRYVRRCKPLKEFM 240
XX
XX 403 LCHDEHEKLFDLVRMLEYDPTQRTTDEALQHPFDLLKK 445
XX 241 LCHDEHEKLFDLVRMLEYDPTQRTTDEALQHPFDLLKK 283
XX
XX
XX RESULT 9
XX AAB58831
XX ID AAB58831 standard; Protein; 350 AA.
XX
XX AAB58831;
XX
XX 27-MAR-2001 (first entry)
XX
XX Breast and ovarian cancer associated antigen protein sequence SEQ ID 539.
XX
XX Human; breast cancer; ovarian cancer; cytosstatic; immunosuppressive;
XX nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; anticancer; vulnerrary; anticonvulsant;
XX antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease.
XX
XX Homo sapiens.
XX
XX WO200055173-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05881.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
```

PI Rosen CA, Ruben SM;  
XX WPI; 2000-611515/58.  
DR N-PSDB; AAF21734.  
XX  
PT New human breast and ovarian cancer associated gene sequences and the  
PT polypeptides encoded by these genes, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -  
XX  
PS Claim 11; Page 970-971; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
CC associated with breast and ovarian cancer. Included in the invention are  
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
CC isolation and characterisation of the DNA and protein sequences of the  
CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
CC antibacterial; antiinflammatory; antitumor; antiparasitic; anticonvulsant;  
CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
CC particularly breast and ovarian cancer. The nucleic acid sequences,  
CC proteins, agonists and antagonists may also be used in the diagnosis,  
CC prevention and treatment of immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular diseases such as  
CC myocardial ischaemia; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases.

XX Sequence 350 AA;

Query Match 63.3%; Score 1526; DB 21; Length 350;  
Best Local Similarity 86.9%; Pred. No. 4.1e-137;  
Matches 279; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 124 EIVDTLGGAGFGKVECTIDHGMNVAVKIVNGYGRYREARSEIOVLEHINSTDPNSV 183  
DB 28 EIVDTLGGAGFGKVECTIDHGMNVAVKIVNGYGRYREARSEIOVLEHINSTDPNST 87  
QY 184 FRCVOMLEMPDHGHVVCIVFELLGISTYDFIKENFLPRQIHITQMAVQICQINFLNH 243  
DB 88 FRCVOMLEMPDHGHVVCIVFELLGISTYDFIKENFLPRQIHITQMAVQICQINFLNH 147  
QY 244 NKLATTDLPKENTLFFVKSADYVVKYNSKMKEDERTLKNITDKVVDGSAFYDEHSTVY 303  
DB 148 NKLATTDLPKENTLFFVKSADYVVKYNSKMKEDERTLKNITDKVVDGSAFYDEHSTVY 207  
QY 304 TRHYRAPAVIILAGMSQPCDWSIGCIIIEYVLGFTVQTHDSKHEHLLAMERILGPDPQ 363  
DB 208 TRHYRAPAVIILAGMSQPCDWSIGCIIIEYVLGFTVQTHDSKHEHLLAMERILGPDPQ 267  
QY 364 MIOKRRKRYFHNNOLDMDHSSAGRYRRCKPLKEMFLCHDEHEKFLDVRMLEYD 423  
DB 268 MIOKRRKRYFHNNOLDMDHSSAGRYRRCKPLKEMFLCHDEHEKFLDVRMLEYD 327  
QY 424 PTOITLDEALQHPFDLLKK 444  
DB 328 PAKRITLREALKHPFDLLKK 348

RESULT 10

AAB33777 ID AAB33777 standard; Protein; 283 AA.

XX AAB33777;

XX 02-FEB-2001 (first entry)

XX Human secreted protein BLAST search protein SEQ ID NO: 121.

KW Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX WO200056753-A1.

XX 28-SEP-2000.

XX 16-MAR-2000; 2000WO-US06765.

XX 23-MAR-1999; 99US-0126051.

XX 10-DEC-1999; 99US-0169906.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-594570/56.

XX Nucleic acid molecules encoding human secreted proteins, used in  
XX preventing, treating or ameliorating a disorder -

XX Disclosure; Page 399-400; 410pp; English.

CC The invention relates to the isolation of genes AAF59277-C59325 encoding  
CC the human secreted proteins AAB33718-B33764. The sequence is a search  
CC result from a BLASTX homology search. The genes and proteins are useful  
CC for preventing, ameliorating or treating medical conditions, e.g. by  
CC protein or gene therapy. The genes are isolated from a range of human  
CC tissues disclosed in the specification. The nucleic acids, proteins,  
CC antibodies and (ant)agonists are useful in the diagnosis, treatment  
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer, and  
CC other cancers of the adrenal gland, bone, bone marrow, breast,  
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemia; (d) wound  
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
CC and (f) infectious diseases such as viral, bacterial, fungal and  
CC parasitic infections.

XX Sequence 283 AA;

Query Match 62.9%; Score 1517; DB 21; Length 283;  
Best Local Similarity 98.6%; Pred. No. 2.2e-136;  
Matches 279; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 163 EAARSEIOVLEHINSTDPNSVFRQVOMLEMPDHGHVVCIVFELLGISTYDFIKENFLP 222  
DB 1 EAARSEIOVLEHINSTDPNSVFRQVOMLEMPDHGHVVCIVFELLGISTYDFIKENFLP 60  
QY 223 QIDHROMAYOICQINFLHNKLTHTDLPKENTLFFVKSADYVVKYNSKMKEDERTLKNITD 282  
DB 61 QIDHROMAYOICQINFLHNKLTHTDLPKENTLFFVKSADYVVKYNSKMKEDERTLKNITD 120  
QY 283 IKVVDGSAFYDEHSTVSTRHYRAPAVIILAGMSQPCDWSIGCIIIEYVLGFTVQ 342  
DB 121 IKVVDGSAFYDEHSTVSTRHYRAPAVIILAGMSQPCDWSIGCIIIEYVLGFTVQ 180  
QY 343 THDSKEHLLAMERILGPDPQIMIOKRRKRYFHNNOLDMDHSSAGRYRRCKPLKEMFL 402  
DB 181 THDSKEHLLAMERILGPDPQIMIOKRRKRYFHNNOLDMDHSSAGRYRRCKPLKEMFL 240  
QY 403 LCHDEHEKFLDVRMLEYDPTORTLDEALQHPFDLLKK 445  
DB 241 LCHDEHEKFLDVRMLEYDPTORTLDEALQHPFDLLKK 283

```
RESULT 11
AAE20171
ID AAE20171 standard; Protein: 352 AA.
XX
AC AAE20171;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human kinase protein consensus.
XX
KW Human; kinase protein; enzyme; cytostatic; osteopathic; gene expression;
KW colon-moderately differentiated adenocarcinoma; chromosome mapping; gene;
KW drug screening; therapeutic; gene therapy; tissue typing; chromosome 5;
KW bone osteosarcoma.
XX
OS Homo sapiens.
XX
PN W0200216567-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26389.
XX
PR 24-AUG-2000; 2000US-227470P.
PR 19-MAR-2001; 2001US-0810671.
XX
PA (APPL-) APPLERA CORP.
XX
PI Yan C, Ye J, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR WPI; 2002-269354/31.
XX
PT New human kinase proteins and nucleic acids, useful in drug screening
PT assays, identifying modulators of kinase activity or treating disorders
PT characterized by absence or unwanted expression of the protein
XX
PS Disclosure; Fig 2; 81pp; English.
XX
CC The invention relates to isolated human kinase proteins and nucleic
CC acids. The nucleic acid and peptide sequences can be used as models for
CC the development of human therapeutic targets, aid in the identification
CC of therapeutic proteins and serve as targets for the development of
CC human therapeutic agents that modulate kinase activity in cells and
CC tissues that express the kinase. The nucleic acids are useful as probes
CC or primers, in constructing recombinant vectors, for expressing
CC antigenic portions of the proteins, chromosome mapping, drug screening,
CC testing an individual for a genotype, and for gene therapy in patients
CC containing cells that are aberrant in kinase gene expression. The
CC proteins may be used in drug screening assays, in the identification of
CC compounds that modulate, stimulate or inhibit kinase activity, in
CC pharmacogenomic analysis, in treating disorders characterized by an
CC absence or unwanted expression of the protein (bone osteosarcoma, or
CC colon-moderately differentiated adenocarcinoma), and in generating
CC antibodies specific for the peptides. Such antibodies can be used to
CC detect the protein in situ, in vitro, or in cell lysate or supernatant,
CC to isolate and purify the proteins from host cells, pharmacogenomic
CC analysis, tissue typing, and in inhibiting protein function. The present
CC sequence is human kinase protein consensus sequence.
XX
SQ Sequence 352 AA;
XX
Query Match 58 1%; Score 1400.5; DB 23; Length 352;
Best Local Similarity 78.0%; Pred.No.4.1e-125; Indels 75; Gaps 37;
Matches 333; Conservative 5; Mismatches 14;
XX
QY 18 SHYLEARSLNERDYRDYRVDYRNDYCEGYVPRHYRDIESGYRIHCKSSVRSRSP 77
DB 1 SHYLER---SNEY--RRY-DEYRNDYGP-----HRD-ES---YHKS---RSLNS 38
XX
QY 78 KKKRNRHCSHOSRSKSHRKRKRSRSTEDDEBGLIQSGDGVLRARYEIVDTIGEGAFGV 137
DB 39 SKKRSH-----SKSHRRKR-RS-EDDEBGLICQSGDVL-ARYEIVDTIGEGAFGV 88
```

```
QY 138 VECIDHGMGMHVAVKIYKRVGRYREARSEIOVLEHINSTDPSVRCYQMLEWPHHG 197
DB 89 VECIDHG---HVAVKIVKRV-RY-EARSEIOVLEHIN-TDPNS-FRCVQMLEWF-HHG 139
QY 198 HVCIVFELAGLSTYDFIKENSFLPQIDHIROMAYQICQSLNFLHNKLTHTDCKPENIL 257
DB 140 H-CIVFELAGLSTYDFIKEN-FLPF--DHIR-MAYQICS--NFL-HNKLTHTDCKPENIL 191
QY 258 FVKSQVYVVKYNSKDKEDERTLKNYDIKVDYFGSATYDDEHSTLVSTRHAPRPAVILALG 317
DB 192 FV-SDY---YNK--KDERLT--NDIKVDYFGSATYDDEHSTLVSTRHAPRPAVILALG 243
QY 318 WSQPCDVMASIGCLILYLGFTVPQTHDSKEHLAMMERILGPIQHMIOKTRKRYPHN 377
DB 244 WSQPCDVMASIGCLILYLGFTV--THDSKEHLAMMERILG--PFHMIOKTRKRYPHN- 299
QY 378 QLPWDEHSSAGRYVRRCKPLKPEMLCHDEHKLPELVARMLEYDPTQITIDEALQHP 437
DB 300 -LPMDEHSSAGRYV--RCKPLKEPML--DEHE-LFPL--MLEYDP--RITL-EAL-HP 345
QY 438 FFDLKK 444
DB 346 FFDLKK 352
XX
RESULT 12
AAW49912
ID AAW49912 standard; Protein: 497 AA.
XX
AC AAW49912;
XX
DT 20-JUL-1998 (first entry)
XX
DE Mouse CLK serine/threonine kinase mCLK2.
XX
KW mCLK2; CLK; serine/threonine kinase; protein kinase; LAMMER kinase;
KW signal transduction; cancer; contraceptive; mouse; therapy;
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Domain 28..45 /note="nuclear localisation domain"
FT FT 163..479
FT Domain /note="catalytic domain"
FT Peptide 388..393 /note="LAMMER motif"
XX
PN W09748723-A2.
XX
PD 24-DEC-1997.
XX
PF 17-JUN-1997; 97WO-IB00946.
XX
PR 19-DEC-1996; 96US-0034286.
PR 17-JUN-1996; 96US-0019629.
PR 09-AUG-1996; 96US-0023485.
PR 13-NOV-1996; 96US-0030860.
PR 15-NOV-1996; 96US-0030964.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Aoki N, Chen Z, Kharitonkov AI, Kim YM, Nayler O;
PI Ulrich A, Wang HY;
DR WPI; 1998-120302/11.
XX
QY New phosphatase and kinase enzyme(s) - useful in the diagnosis and
PT treatment of signal transduction disorders
XX
PS Claim 11; Fig 4; 138pp; English.
```



CC This polypeptide comprises novel mouse CLK serine/threonine kinase  
CC mCLK2, from the CLK serine/threonine kinase family of proteins that  
CC regulate RNA splicing in cells. mCLK2 cDNA was cloned from a mouse  
CC embryo 11.5 p.c. 12AP cDNA library. The invention discloses the use  
CC of discovery of novel protein kinases mCLK2, mCLK3 (see AAM49912) and  
CC mCLK4 (see AAM49914) of mol.wt. 59.9, 58.5 and 57.2 kDa,  
CC respectively, as well as other novel proteins (see AAM49906-10)  
CC involved in cellular signal transduction, and provides vectors,  
CC host cells, purified recombinant proteins, methods for identifying  
CC compounds that activate or inhibit the novel proteins, as well as  
CC methods for the diagnosis and treatment of diseases associated with  
CC the novel proteins. Overexpression of CLK serine/threonine kinases  
CC has been implicated in certain types of cancer. Compounds that  
CC inhibit their catalytic activity or disrupt their interactions  
CC with natural binding partners may act as anti-cancer therapeutics.  
CC mCLK related molecules and compounds may also be useful as male  
CC contraceptives.

XX Sequence 497 AA;

Query Match 55.2%; Score 1311.5; DB 19; Length 497;  
Best Local Similarity 59.0%; Pred. No. 2.7e-118;

Matches 259; Conservative 62; Mismatches 105; Indels 13; Gaps 8;

QY 13 EDTFHSHTLEARS-----LNEFDYDRRYVDEY-RNDYCEGYVPRHYHDISSGYRINCS 66  
DB 47 EDSR--YHVRSSSYDDHSDRLYDRYCGSYRNDYGRDGEAYVDTDFQSYEYHRE 104  
QY 67 KSSVSRSSS-PKRENHRCSSHQSRKSRHRRKRSITDDEGHILCOGQVLAARYI 125  
DB 105 NSSYSQSSSRKRRRRRRRTFSRSSHSSRRKASVEDDAGHLIYHVGWMLDERYEI 164  
QY 126 VDTLEGAFGKVEICIDHGMGMHVAVKIKVNGVREARASEIOVLEHINSTDPNSVFR 185  
DB 165 VSTLEGGSGRNVGICIDRVRGRRLVKIKVVEKKEARLEINVLEKINCKDRKNTL 224  
QY 186 CYQMLEWPDHGHVCIPELLGLSTYDPKENSFLPEFOIDHTRQMAVOICQGINFLHNK 245  
DB 225 CYQMDWPDYHGMICISPELLGLSTPDKNNYLPYPIHQVRHMAFOLCAVKEFLHDK 284  
QY 246 LHTDLPKPNILFVSDYVVKYNSMKKRDERT-LKNTDICKVDPGSATDDEHSTIVST 304  
DB 285 LHTDLPKPNILFVNSDYELT-NPLEKRDERTSVKSTAR-VDPGSATFDHHSSTLST 342  
QY 305 RHVRAPEVILALGWSQPCDWSIGCILLEYVLGFTVFOHDSKEHLAMMERILGPIPOHM 364  
DB 343 RHVRAPEVILALGWSQPCDWSIGCIFIEYVLGFLVQPTHNSRHLAMMERILGVPSPRM 402  
QY 365 IOKTKRKYFHHNOIDMBHSAGRYVRRCKPLKEFMILCHDEHEKFLDYLRMLEYDP 424  
DB 403 IRTKRQKYFYGRIDMBENTSAGRYVRENCKPLERY-LTSEADHQLFDLIENMLEYEP 461  
QY 425 TORITLDEALQHPFFDLK 443  
DB 462 AKRLTLGHALQHPFACLR 480

RESULT 13  
ABG23354  
ID ABG23354 standard; Protein; 490 AA.  
XX  
AC ABG23354;

DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #23345.  
XX

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX Food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
XX  
PN MO200175067-AA.

XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001MO-US08631.  
PF  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR  
XX 23-AUG-2000; 2000US-0649167.  
PA  
(HYSE-) HYSEQ INC.  
PI  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI, 2001-639362/73.  
DR  
N-PSDB; AAS87541.  
PT  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 20; SEQ ID No 53713; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 490 AA;

Query Match 54.5%; Score 1314.5; DB 22; Length 490;  
Best Local Similarity 59.8%; Pred. No. 1.1e-116;  
Matches 251; Conservative 59; Mismatches 89; Indels 21; Gaps 7;

QY 34 RRYVDE--YRNDYCEGYVPR-HYHDISSGYRINCSKSVSRSSPKRKR-----NR 83  
DB 64 RRYHSSERGSRGSYEHNRSRKHQRSSRS-----WSSSDSDTRR---RRRDSYHVR 114  
QY 84 HCSSHQSRKSK-HRRKRSRISIDDEEGHILCOGQVLAARYIYVDTLEGAFGKVEICID 142  
DB 115 RCSRFTSSRSSQSHSSRKAKSYEDDTEGHILYHVGWMLDERYIVSTLGKTFGRVVOQVD 174  
QY 143 HGMGMHVAVKIKVNGVREARASEIOVLEHINSTDPNSVRCYQMLEWPDHGHVCI 202  
DB 175 HRRRGARVALKTIKVEYKKEARLEIVLEKINCKDQKNL-CYQMDWPDYHGMICIS 233  
QY 203 FELLGLSTYDFTIKENSFLPEFOIDHTRQMAVOICQGINFLHNKLTHTDLPKPNILFVKS 262  
DB 234 LELGLSTPFDPLKONNHLPYPIHQVHMAVSQCAVKEFLHDKLHTDLPKPNILFVNSD 293  
QY 263 YVVKYNSCMKGRDERTLKNTDICKVDPGSATYDDEHSHSTVSTRHRAPEVILALGWSQPC 322  
DB 294 YELTYNLKKKHERSVKSTAVKVGDFGSATPEHEHSHSTVSTRHRAPEVILALGWSQPC 353  
QY 323 DVMSIGCLLIEYVLGFTVFOHDSKEHLAMMERILGPIPOHVIKOTRKRKYFHHNOIDMD 382  
DB 354 DVMSIGCLLIEYVGVFTLFOHDKRQHLATMERILGPIPSRIRTKRKQKYFYGRGLWD 413

QY 383 EHSAGYVRRCKPLKEFMLCHDEHEKLFDLVRMLFYDPTORITLDEALQHPFDLL 442  
 DB 414 ENTSAGRYVRNCKPLKQVLTSEAEEDHQFLDLESMLERYPAQRLTLGALQHPFSRL 473

RESULT 14  
 ID AAB85506 standard; protein; 431 AA.  
 AC AAB85506;  
 DT 25-SEP-2001 (first entry)  
 XX  
 DE Human protein kinase SGK090.  
 XX  
 KW Protein kinase; enzyme; cytosolic; noctropic; neuroprotective; human;  
 KW antiParkinsonian; virucide; antibacterial; antifungal; antimigraine;  
 KW analgesic; hypotensive; hypertensive; immunosuppressive; antiallergic;  
 KW antispasmodic; antirheumatic; antiarthritic; ophthalmological; anorectic;  
 KW osteopathic; thrombolytic; antiarteriosclerotic; antiaschematic;  
 KW vasotropic; antidiabetic; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 27 /note= "encoded by CAN"  
 FT Misc-difference 28 /note= "encoded by NAG"  
 FT  
 PN MO200155356-A2.  
 PD 02-AUG-2001.  
 PF 25-JAN-2001; 2001WO-US02337.  
 XX  
 PR 25-JAN-2000; 2000US-0178078.  
 PR 31-JAN-2000; 2000US-0179364.  
 PR 17-FEB-2000; 2000US-0183173.  
 PR 17-MAR-2000; 2000US-0190162.  
 PR 29-MAR-2000; 2000US-0193404.  
 PR 13-NOV-2000; 2000US-0247013.  
 XX  
 PA (SUGEN-) SUGEN INC.  
 XX  
 PI Plowman G, Whyte D, Manning G, Sudarsanam S, Martinez R;  
 XX  
 DR WPI; 2001-476202/51.  
 DR N-PSDB; AAB46906.  
 XX  
 PT Kinase polypeptides useful for treating cancers, Alzheimer's disease,  
 PT viral infections, diabetes, obesity, organ transplant rejection and  
 PT rheumatoid arthritis -  
 XX  
 PS Claim 7; Page 216; 218bp; English.  
 XX  
 CC The invention provides human protein kinases and protein kinase-like  
 CC enzymes and polynucleotides encoding the polypeptides. The kinase  
 CC polypeptides and their modulators are useful for treating a disease or  
 CC disorder such as cancer, immune-related diseases, cardiovascular disease,  
 CC brain or neuronal-associated disease and metabolic disorders, including  
 CC cancers of tissues, cancers of hematopoietic origin, diseases of the  
 CC central nervous system, diseases of the peripheral nervous system,  
 CC Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic  
 CC lateral sclerosis, viral infections, infections caused by prions,  
 CC bacteria and fungi, ocular diseases, migraines, pain, sexual dysfunction,  
 CC mood disorders, attention disorders, cognitive disorders, hypotension,  
 CC hypertension, psychotic disorders, neurological disorders, dyskinesias,  
 CC metabolic disorders, and organ transplant rejection. They are also useful  
 CC for treating rhinitis, autoimmunity, atherosclerosis, psoriasis,  
 CC osteoarthritis, asthma, chronic inflammatory pelvic disease, chronic  
 CC inflammatory bowel disease, rheumatoid arthritis, metabolic disorders

CC such as diabetes, obesity, cardiovascular diseases such as reperfusion  
 CC injury, coronary thrombosis, clotting disorders and atherosclerosis,  
 CC ocular diseases such as glaucoma, retinopathy and macular degeneration,  
 CC psychiatric and neurological disorders such as anxiety, schizophrenia,  
 CC dementia, manic depression, etc. The polynucleotides are useful in gene  
 CC therapy techniques to treat the above mentioned disorders. Sequences  
 CC AAB85491-85522 represent the human protein kinases of the invention.  
 XX  
 SQ Sequence 431 AA;  
 Query Match 54.4%; Score 1310.5; DB 22; Length 431;  
 Best Local Similarity 59.2%; Pred. No 2.2e-116;  
 Matches 248; Conservative 61; Mismatches 91; Indels 19; Gaps 6;

QY 34 RRYVDE--YRNDYCEGYVPRHYHRDLESQYRIHCKSVSRSSPKRRK-----NRH 84  
 DB 5 RRYHSSBRGSRGSGYCEHYRSRCKXQRRSRSM---SSSDRTRR-----RRRDSYHVR 56  
 QY 85 CSHQSSKSK-HRRKSRSLIEDDEBGLICQSGDVLPRRIEIVDTLGBAGFKYVECTIDH 143  
 DB 57 CSRFTFSRSSQSHSRKAKSVYEDDTGHLIYHVDWLQERYEIVSTLKGTFGRVVCVDH 116  
 QY 144 GMDQMVAVKIVXNVGRYREARSEIQVLEHLNSTDPNSVPRCYQMLEWFDHGHVCIVF 203  
 DB 117 RRRGARVALKIIQVKEKKAABLEIKYLEKINEKDPGRVL-CYQMDFWDHGMICSL 175  
 QY 204 ELIGLSTYDFIKENSFLPFOIDIRQWAVQICQINFLHNKLTHDLKRENTILFYKSDY 263  
 DB 176 ELIGLSTFDPLKDNHLPYPIHQVHMAQLOQAVKFLHDKLTHDLKRENTILFVNSDY 235  
 QY 264 VVKYNSKMKRDEFTLKNITDKVYDFGSATYDDDEHSHLTVSTRYRAPEVILALGMSQPCD 323  
 DB 236 ELTYNLEKRRHESVSTARVDFGSATYDDDEHSHLTVSTRYRAPEVILALGMSQPCD 295  
 QY 324 VMSIGCLILEYVYGFVFTDHSKEHLAMMERITLGPQMIQKTRKRYFHHNQMDSE 383  
 DB 296 VMSIGCLIPRYVYGFVFTDHSKEHLAMMERITLGPQMIQKTRKRYFHHNQMDSE 355  
 QY 384 HSSAGYVRRCKPLKEFMLCHDEHEKLFDLVRMLFYDPTORITLDEALQHPFDLL 442  
 DB 356 NTSAGRYVRNCKPLKQVLTSEAEEDHQFLDLESMLERYPAQRLTLGALQHPFSRL 414

RESULT 15  
 ID AAM49913 standard; Protein; 485 AA.  
 AC AAM49913;  
 XX  
 DT 20-JUL-1998 (first entry)  
 XX  
 DE Mouse CLK serine/threonine kinase mCLK3.  
 XX  
 KW mCLK3; CLK; serine/threonine kinase; protein kinase; LAMMER kinase;  
 KW signal transduction; cancer; contraceptive; mouse; therapy;  
 KW diagnosis.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 23..41  
 FT Domain /note= "nuclear localisation domain"  
 FT Domain 151..469  
 FT Peptide /note= "catalytic domain"  
 FT Peptide 377..382  
 FT Peptide /note= "LAMMER motif"  
 PN MO9748723-A2.  
 PD 24-DEC-1997.  
 XX  
 PF 17-JUN-1997; 97WO-IB00946.  
 XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 15:16:24 ; Search time 21 Seconds  
(Without alignments)  
896.587 Million cell updates/sec

Title: US-10-339-656-2

Perfect score: 2410  
Sequence: 1 MCIPLEASHVEDTHPSHY.....QRTIDRALQHPDILKKK 445

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCBUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2410	100.0	445	4 US-09-810-671-2	Sequence 2, Appl1
2	2312	95.9	427	4 US-09-810-671-4	Sequence 4, Appl1
3	2117.5	87.9	451	2 US-09-016-000-3	Sequence 3, Appl1
4	1887	78.3	429	4 US-09-810-671-5	Sequence 5, Appl1
5	1887	78.3	484	4 US-09-457-040B-12	Sequence 12, Appl1
6	1417	58.8	499	4 US-09-457-040B-13	Sequence 13, Appl1
7	1222	50.7	517	4 US-09-457-040B-14	Sequence 14, Appl1
8	695	28.8	736	4 US-09-457-040B-14	Sequence 26, Appl1
9	550.5	22.8	527	4 US-09-659-166-2	Sequence 2, Appl1
10	546.5	22.7	528	2 US-08-802-466-2	Sequence 2, Appl1
11	546.5	22.7	528	2 US-09-350-484-2	Sequence 2, Appl1
12	544.5	22.6	568	2 US-08-835-170-4	Sequence 4, Appl1
13	544.5	22.6	568	3 US-09-359-257-4	Sequence 4, Appl1
14	544.5	22.6	568	4 US-09-371-674-4	Sequence 4, Appl1
15	544.5	22.6	588	2 US-08-835-170-2	Sequence 2, Appl1
16	544.5	22.6	588	2 US-09-359-257-2	Sequence 2, Appl1
17	544.5	22.6	588	4 US-09-371-674-2	Sequence 2, Appl1
18	523	21.7	508	2 US-08-818-024-3	Sequence 3, Appl1
19	523	21.7	508	3 US-09-334-775A-3	Sequence 3, Appl1
20	517	21.5	620	3 US-09-126-646-2	Sequence 2, Appl1
21	517	21.5	620	4 US-09-421-491-2	Sequence 2, Appl1
22	505	21.0	539	2 US-08-818-024-4	Sequence 4, Appl1
23	505	21.0	539	3 US-09-334-775A-4	Sequence 4, Appl1
24	505	21.0	539	3 US-08-789-275-6	Sequence 6, Appl1
25	478.5	19.9	557	3 US-09-027-064-2	Sequence 2, Appl1
26	478.5	19.9	557	3 US-09-271-815-2	Sequence 2, Appl1
27	472.5	19.6	763	2 US-08-677-862-2	Sequence 2, Appl1

28	472.5	19.6	763	2 US-09-252-571-2	Sequence 2, Appl1
29	472.5	19.6	763	3 US-09-434-065-2	Sequence 2, Appl1
30	472.5	19.6	763	3 US-08-789-275-4	Sequence 4, Appl1
31	472.5	19.6	763	3 US-08-789-275-5	Sequence 5, Appl1
32	419	17.4	1209	4 US-09-749-588-4	Sequence 4, Appl1
33	407	16.9	1170	4 US-09-749-588-2	Sequence 4, Appl1
34	392	16.3	1087	1 US-08-264-002-5	Sequence 5, Appl1
35	359.5	14.9	376	2 US-08-818-024-1	Sequence 1, Appl1
36	359.5	14.9	379	2 US-09-334-775A-1	Sequence 1, Appl1
37	343	14.2	353	1 US-08-176-620A-14	Sequence 14, Appl1
38	343	14.2	353	4 US-08-461-985-14	Sequence 14, Appl1
39	332.5	13.8	607	4 US-09-417-197-47	Sequence 47, Appl1
40	331.5	13.8	655	1 US-08-264-002-2	Sequence 2, Appl1
41	328	13.6	379	4 US-09-457-040B-36	Sequence 36, Appl1
42	326.5	13.5	360	1 US-08-674-612-4	Sequence 4, Appl1
43	326.5	13.5	360	1 US-08-469-421-12	Sequence 12, Appl1
44	326.5	13.5	360	1 US-08-469-421-14	Sequence 14, Appl1
45	326.5	13.5	360	1 US-08-250-975-12	Sequence 12, Appl1

ALIGNMENTS

RESULT 1  
US-09-810-671-2  
; Sequence 2, Application US/09810671  
; Patent No. 6455291  
; GENERAL INFORMATION:  
; APPLICANT: YAN Chunhua et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: C1000758  
; CURRENT APPLICATION NUMBER: US/09/810,671  
; CURRENT FILING DATE: 2001-06-08  
; NUMBER OF SEQ. ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ. ID NO 2  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Human  
; US-09-810-671-2

Query Match 100.0%; Score 2410; DB 4; Length 445;  
Best Local Similarity 100.0%; Pred. No. 4, 1e-218;  
Matches 445; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MCIPLEASHVEDTHPSHYLEARSINERDYRDYVDEYRNDYCEGYVPRHYRDI	EG 60
DB	1	MCIPLEASHVEDTHPSHYLEARSINERDYRDYVDEYRNDYCEGYVPRHYRDI	EG 60
QY	61	YRHCSSVRSRSPKRRKRRHSHSRSSKSHRKRSTEDDEEHLICQSGVLR	120
DB	61	YRHCSSVRSRSPKRRKRRHSHSRSSKSHRKRSTEDDEEHLICQSGVLR	120
QY	121	ARVEYDTLGEAGFGKVEICIDHGMGMHVAVKIVNNGVRRARASEIOVL	180
DB	121	ARVEYDTLGEAGFGKVEICIDHGMGMHVAVKIVNNGVRRARASEIOVL	180
QY	181	NSFRVCQMLEWDDHGHVCIYVELLGLSTYDFIKNSFLPOIDIRHMAVOI	240
DB	181	NSFRVCQMLEWDDHGHVCIYVELLGLSTYDFIKNSFLPOIDIRHMAVOI	240
QY	241	LHNKLTHTDLPENILFYKSPDVVYKSKMRDETLKNTDKVVDGSGATYD	300
DB	241	LHNKLTHTDLPENILFYKSPDVVYKSKMRDETLKNTDKVVDGSGATYD	300
QY	301	LVSTRHRAPEVIALGWSQPCDWSIGCILLEYVIGFTVFTQTHDSKEHL	360
DB	301	LVSTRHRAPEVIALGWSQPCDWSIGCILLEYVIGFTVFTQTHDSKEHL	360
QY	361	POHMTOKTKKRYFFHNQJLDWDEHSAGYVRRRCPLKEPMKCHDEEBK	420
DB	361	POHMTOKTKKRYFFHNQJLDWDEHSAGYVRRRCPLKEPMKCHDEEBK	420

Db 361 PQHMOIKTRKKYFHHNQLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLYRML 420  
QY 421 EYDPTQRIITLDEALQHPFDLLKKK 445  
Db 421 EYDPTQRIITLDEALQHPFDLLKKK 445

## RESULT 2

US-09-810-671-4  
; Sequence 4, Application US/09810671  
; Patent No. 6455291  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CLO00758  
; CURRENT APPLICATION NUMBER: US/09/810,671  
; CURRENT FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 427  
; TYPE: PRT  
; ORGANISM: Human  
US-09-810-671-4

Query Match 95.9%; Score 2312; DB 4; Length 427;  
Best Local Similarity 100.0%; Pred. No. 6.1e-209;  
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 HYLEARSINERDVRDRYVDEYNDYCEGVPRHHYRDITSGYRHCSSKSVSRSSPK 78  
Db 1 HYLEARSINERDVRDRYVDEYNDYCEGVPRHHYRDITSGYRHCSSKSVSRSSPK 60  
QY 79 RKNRHCSSHOSKSHRRKRSRSIEDDEBGLICSGDYLARAYETVDTLGGARQKV 138  
Db 61 RKNRHCSSHOSKSHRRKRSRSIEDDEBGLICSGDYLARAYETVDTLGGARQKV 120  
QY 139 ECTIDHMDGMHVAVKIVKNGRYRRAARSEIQVLEHINSTDPNSVRCVOMLEWPDHGH 198  
Db 121 ECTIDHMDGMHVAVKIVKNGRYRRAARSEIQVLEHINSTDPNSVRCVOMLEWPDHGH 180  
QY 199 VCIVFELGLSTYDFIKENSFLPFOIDHROMAYOICQGINFLHNKLTHTDLKPNILF 258  
Db 181 VCIVFELGLSTYDFIKENSFLPFOIDHROMAYOICQGINFLHNKLTHTDLKPNILF 240  
QY 259 VKSDYVVKYNSKMKRDERTLKNTDIKVDYFGSATYDDEHSTLVSTRHYRAPEVILALGM 318  
Db 241 VKSDYVVKYNSKMKRDERTLKNTDIKVDYFGSATYDDEHSTLVSTRHYRAPEVILALGM 300  
QY 319 SPPCDVMSIGCTILIEYLGFTVQTHDSKEHLAMMERILGPIPOHMIQTRKKRYFHNO 378  
Db 301 SPPCDVMSIGCTILIEYLGFTVQTHDSKEHLAMMERILGPIPOHMIQTRKKRYFHNO 360  
QY 379 LDMDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLYRMLBYDPTQRIITLDEALQHPF 438  
Db 361 LDMDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLYRMLBYDPTQRIITLDEALQHPF 420  
QY 439 FDLKKK 445  
Db 421 FDLKKK 427

## RESULT 3

US-09-016-000-3  
; Sequence 3, Application US/09016000  
; Patent No. 5962232  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Ial, Preeti  
; APPLICANT: Bandman, Olga  
; APPLICANT: Akerdion, Ingrid E.

APPLICANT: Shah, Purvi  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl G.  
TITLE OF INVENTION: PROTEIN KINASE MOLECULES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,000  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0465 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 451 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: NEUTPMT01  
CLONE: 339963  
US-09-016-000-3

Query Match 87.9%; Score 2117.5; DB 2; Length 451;  
Best Local Similarity 88.9%; Pred. No. 1.2e-190;  
Matches 399; Conservative 3; Mismatches 6; Indels 41; Gaps 3;

QY 8 SH-SVEEDTH-----PSHYLEARSINERDVRDRYVDEYNDYCEGVPRHHYRD 56  
Db 33 SHSSTQENRCHCKPHQFKESDCHYLEARSINERDVRDRYVDEYNDYCEGVPRHHYRD 92  
QY 57 ISSGYRHCSSKSVSRSSPKRKNRHCSSHOSKSHRRKRSRSIEDDEBGLICQSG 116  
Db 93 ISSGYRHCSSKSVSRSSPKRKNRHCSSHOSKSHRRKRSRSIEDDEBGLICQSG 128  
QY 117 DYLARAYETVDTLGGARQKVVECTIDHMDGMHVAVKIVKNGRYRRAARSEIQVLEHIN 176  
Db 129 -----MKSVDTLGGARQKVVECTIDHMDGMHVAVKIVKNGRYRRAARSEIQVLEHIN 182  
QY 177 STDPSVRCVOMLEWPDHGHVCTIVFELGLSTYDFIKENSFLPFOIDHROMAYOICQ 236  
Db 183 STDPSVRCVOMLEWPDHGHVCTIVFELGLSTYDFIKENSFLPFOIDHROMAYOICQ 242  
QY 237 SINFLHNKLTHTDLKPNILFVKSDDYVVKYNSKMKRDERTLKNTDIKVDYFGSATYDDE 296  
Db 243 SINFLHNKLTHTDLKPNILFVKSDDYVVKYNSKMKRDERTLKNTDIKVDYFGSATYDDE 302  
QY 297 HHSTLVSTRHYRAPEVILALGMSQPCDVMSIGCTILIEYLGFTVQTHDSKEHLAMMERI 356  
Db 303 HHSTLVSTRHYRAPEVILALGMSQPCDVMSIGCTILIEYLGFTVQTHDSKEHLAMMERI 362  
QY 357 LGPIPOHMIQTRKKRYFHHNQLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLY 416

Db 363 LGPIPOHMIQTRKRRKYFHHNQLDWDDEHSSAGRYVRRCCKPLKEFMLCHDEHEKTLFDLV 422  
 QY 417 RRMLEYDPTORTITLDEALQHPFDLKK 445  
 Db 423 RRMLEYDPTORTITLDEALQHPFDLKK 451

RESULT 4  
 US-09-810-671-5  
 ; Sequence 5, Application US/09810671  
 ; Patent No. 6455291  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAN, Chunhua et al  
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
 ; FILE REFERENCE: C1000758  
 ; CURRENT APPLICATION NUMBER: US/09/810,671  
 ; CURRENT FILING DATE: 2001-06-08  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: FASTSEQ for Windows Version 4.0  
 ; SEQ ID NO 5  
 ; LENGTH: 429  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-810-671-5

Query Match 78.3%; Score 1887; DB 4; Length 429;  
 Best Local Similarity 82.1%; Pred. No. 5, 2e-169;  
 Matches 352; Conservative 30; Mismatches 45; Indels 2; Gaps 2;

QY 18 SHYLEASLNEDRYDRRYVDEYNDYCEGYPRHYHNDIESGRIHCSKSVRSRSP 77  
 Db 1 SHYLEASLNEDRYDRRYVDEYNDYCEGYPRHYHNDIESGRIHCSKSVRSRSP 60  
 QY 78 KRR-RNRKCSH-QSRKSHRRKRSRSIEDDEEGLICQSGDVLARAYEIVDTLGGAGF 135  
 Db 61 KSKRIHSHSTRSHSGHSHRRKRTSRVEDDEEGLICQSGDVLARAYEIVDTLGGAGF 120  
 QY 136 KVECIDHGMQDMHVAVKIVKVGRYREARSEIOVLEHNTDPSVFRVCQMLEFSDH 195  
 Db 121 KVECIDHKGAGRHVAVKIVKVGDRYCEARSEIOVLEHNTDPSVFRVCQMLEFSDH 180  
 QY 196 HGHVCIFFELGLSTYDFIKENSFLPFOIDIRQMAVOICQINFLHNKLTHTDLKPEN 255  
 Db 181 HGHVCIFFELGLSTYDFIKENSFLPFRDHIRKMAVOICQINFLHNKLTHTDLKPEN 240  
 QY 256 ILFVKSADYVVKNSKMKRDERLTAKNTDIKVVDFGSATYDDEHSTLVSTRHYRAPEVILA 315  
 Db 241 ILFVQSDYTAAYNPKIKRDERLTAKNTDIKVVDFGSATYDDEHSTLVSTRHYRAPEVILA 300  
 QY 316 LGMSQPCDVMSIGCIIIEYLGFTVFQTHDSKEHLAMMERLIGPIPOHMIQTRKRYFH 375  
 Db 301 LGMSQPCDVMSIGCIIIEYLGFTVFQTHDSKEHLAMMERLIGPIPOHMIQTRKRYFH 360  
 QY 376 HNOIDWDEHSSAGRYVRRCCKPLKEFMLCHDEHEKTLFDLVRLMEYDPTORTITLDEALQ 435  
 Db 361 HNOIDWDEHSSAGRYVRRCCKPLKEFMLSQDVEHERLFDLQKLEIDPAKRTITLDEALQ 420  
 QY 436 HPFDLKK 444  
 Db 421 HPFDLKK 429

RESULT 5  
 US-09-457-040B-12  
 ; Sequence 12, Application US/09457040B  
 ; Patent No. 6387641  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Vertex Pharmaceuticals Incorporated  
 ; APPLICANT: Bellon, Steve  
 ; TITLE OF INVENTION: Crystallized P38 Complexes  
 ; FILE REFERENCE: VPI/98-14

; CURRENT APPLICATION NUMBER: US/09/457,040B  
 ; CURRENT FILING DATE: 1999-12-08  
 ; NUMBER OF SEQ ID NOS: 41  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 12  
 ; LENGTH: 484  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-457-040B-12

Query Match 78.3%; Score 1887; DB 4; Length 484;  
 Best Local Similarity 82.1%; Pred. No. 6, 2e-169;  
 Matches 352; Conservative 30; Mismatches 45; Indels 2; Gaps 2;

QY 18 SHYLEASLNEDRYDRRYVDEYNDYCEGYPRHYHNDIESGRIHCSKSVRSRSP 77  
 Db 54 SHYLEASLNEDRYDRRYVDEYNDYCEGYPRHYHNDIESGRIHCSKSVRSRSP 113  
 QY 78 KRR-RNRKCSH-QSRKSHRRKRSRSIEDDEEGLICQSGDVLARAYEIVDTLGGAGF 135  
 Db 114 KSKRIHSHSTRSHSGHSHRRKRTSRVEDDEEGLICQSGDVLARAYEIVDTLGGAGF 173  
 QY 136 KVECIDHGMQDMHVAVKIVKVGRYREARSEIOVLEHNTDPSVFRVCQMLEFSDH 195  
 Db 174 KVECIDHKGAGRHVAVKIVKVGDRYCEARSEIOVLEHNTDPSVFRVCQMLEFSDH 233  
 QY 196 HGHVCIFFELGLSTYDFIKENSFLPFOIDIRQMAVOICQINFLHNKLTHTDLKPEN 255  
 Db 234 HGHVCIFFELGLSTYDFIKENSFLPFRDHIRKMAVOICQINFLHNKLTHTDLKPEN 293  
 QY 256 ILFVKSADYVVKNSKMKRDERLTAKNTDIKVVDFGSATYDDEHSTLVSTRHYRAPEVILA 315  
 Db 294 ILFVQSDYTAAYNPKIKRDERLTAKNTDIKVVDFGSATYDDEHSTLVSTRHYRAPEVILA 353  
 QY 316 LGMSQPCDVMSIGCIIIEYLGFTVFQTHDSKEHLAMMERLIGPIPOHMIQTRKRYFH 375  
 Db 354 LGMSQPCDVMSIGCIIIEYLGFTVFQTHDSKEHLAMMERLIGPIPOHMIQTRKRYFH 413  
 QY 376 HNOIDWDEHSSAGRYVRRCCKPLKEFMLCHDEHEKTLFDLVRLMEYDPTORTITLDEALQ 435  
 Db 414 HNOIDWDEHSSAGRYVRRCCKPLKEFMLSQDVEHERLFDLQKLEIDPAKRTITLDEALQ 473  
 QY 436 HPFDLKK 444  
 Db 474 HPFDLKK 482

RESULT 6  
 US-09-457-040B-13  
 ; Sequence 13, Application US/09457040B  
 ; Patent No. 6387641  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Vertex Pharmaceuticals Incorporated  
 ; APPLICANT: Bellon, Steve  
 ; TITLE OF INVENTION: Crystallized P38 Complexes  
 ; FILE REFERENCE: VPI/98-14  
 ; CURRENT APPLICATION NUMBER: US/09/457,040B  
 ; CURRENT FILING DATE: 1999-12-08  
 ; NUMBER OF SEQ ID NOS: 41  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 13  
 ; LENGTH: 499  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-457-040B-13

Query Match 58.8%; Score 1417; DB 4; Length 499;  
 Best Local Similarity 59.7%; Pred. No. 9, 3e-125;  
 Matches 259; Conservative 69; Mismatches 98; Indels 8; Gaps 3;

QY 18 SHYLEASLNEDRYDRRYVDEYNDYCEGYPRHYHNDIESGRIHCSKSVRSRSP 71  
 Db 50 SHYLEASLNEDRYDRRYVDEYNDYCEGYPRHYHNDIESGRIHCSKSVRSRSP 109





FILE REFERENCE: GP-10218  
CURRENT APPLICATION NUMBER: US/09/659,166  
CURRENT FILING DATE: 2000-09-11  
PRIOR APPLICATION NUMBER: UK 9921505.5  
PRIOR FILING DATE: 1999-09-10  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 527  
TYPE: PRT  
ORGANISM: rattus  
US-09-659-166-2

Query Match 22.8%; Score 550.5; DB 4; Length 527;  
Best Local Similarity 34.3%; Pred. No. 2,5e-43;  
Matches 121; Conservative 69; Mismatches 122; Indels 41; Gaps 7;

QY 105 DDEBGLICQSGDVLARAYIVDTLGGAGFKVVECTDHGMGMHVAVKIKVNGRYREA 164  
DB 130 DDDQGSYYQVPHDHVAAYREVLKVIKGSFGQVYKAYDHKVH-OHVALKMYRNEKRFHRQ 188  
QY 165 ARSEIOVLEHINSTDPNSVFRVOMLEWFDHGHGVCIVFELLGISTYDFIKENSFLPFOI 224  
DB 189 AAEIRILEHLRKQDKNTMNVIMHLENFTFRNHICMTPELLSMNLVYLKIKKFKQGSFL 248  
QY 225 DHIRQAVYQICQINFILHNKLTHTDLPENILFVKSQDYVVKYNSKMKRDERTLKNTDIK 284  
DB 249 PLVKRFASHILQCDLALHKRRIHCDLKPENIL-----LKQGR-----SGIK 291  
QY 285 VDPGSAATYDDEHNSLTAVSTRHARPEVILALGMSQPCDVMSIGCILIEVYLGFTVPOTH 344  
DB 292 VIDFSSCYEHQRYTYIQRFRAPREVILGARVGMPIDMSVLSGCLLAEILLTGYPPLPGE 351  
QY 345 DSKEHLAMMERILGPIPOHMIQKTRKRYFHNNQ-----LDWDEHSSAGRYVVR 393  
DB 352 DRGQDLACMTELGMPSQKLDASRAKNFVSXGPRYCTVTTLSDGSVVLNCGRSRSG 411  
QY 394 RCKPLKE-----FMLCHDEHEKLF-DLYRMLEIYPTQRTITDEALQHPF 438  
DB 412 KLKGPESREMGNALKQCD-----PLFLDPLKQCLEMDPAVRMTFGQALRHPM 460

## RESULT 10

US-08-802-466-2  
Sequence 2, Application US/08802466  
Patent No. 5972606  
GENERAL INFORMATION:  
APPLICANT: Creasy, et al.  
TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,466  
FILING DATE: 19 February 1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: William T. Han  
REGISTRATION NUMBER: 34,344  
REFERENCE/DOCKET NUMBER: GH50002

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5219  
TELEFAX: 610-270-4026  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 528 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-802-466-2

Query Match 22.7%; Score 546.5; DB 2; Length 528;  
Best Local Similarity 34.3%; Pred. No. 6e-43;  
Matches 121; Conservative 68; Mismatches 123; Indels 41; Gaps 7;

QY 105 DDEBGLICQSGDVLARAYIVDTLGGAGFKVVECTDHGMGMHVAVKIKVNGRYREA 164  
DB 131 DDDQGSYYQVPHDHVAAYREVLKVIKGSFGQVYKAYDHKVH-OHVALKMYRNEKRFHRQ 189  
QY 165 ARSEIOVLEHINSTDPNSVFRVOMLEWFDHGHGVCIVFELLGISTYDFIKENSFLPFOI 224  
DB 190 AAEIRILEHLRKQDKNTMNVIMHLENFTFRNHICMTPELLSMNLVYLKIKKFKQGSFL 249  
QY 225 DHIRQAVYQICQINFILHNKLTHTDLPENILFVKSQDYVVKYNSKMKRDERTLKNTDIK 284  
DB 250 PLVKRFASHILQCDLALHKRRIHCDLKPENIL-----LKQGR-----SGIK 292  
QY 285 VDPGSAATYDDEHNSLTAVSTRHARPEVILALGMSQPCDVMSIGCILIEVYLGFTVPOTH 344  
DB 293 VIDFSSCYEHQRYTYIQRFRAPREVILGARVGMPIDMSVLSGCLLAEILLTGYPPLPGE 352  
QY 345 DSKEHLAMMERILGPIPOHMIQKTRKRYFHNNQ-----LDWDEHSSAGRYVVR 393  
DB 353 DRGQDLACMTELGMPSQKLDASRAKNFVSXGPRYCTVTTLSDGSVVLNCGRSRSG 412  
QY 394 RCKPLKE-----FMLCHDEHEKLF-DLYRMLEIYPTQRTITDEALQHPF 438  
DB 413 KLKGPESREMGNALKQCD-----PLFLDPLKQCLEMDPAVRMTFGQALRHPM 461

## RESULT 11

US-09-350-484-2  
Sequence 2, Application US/09350484  
Patent No. 6159716  
GENERAL INFORMATION:  
APPLICANT: Creasy, et al.  
TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/350,484  
FILING DATE: 09-Jul-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/802,466  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: William T. Han  
REGISTRATION NUMBER: 34,344  
REFERENCE/DOCKET NUMBER: GH50002







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK protein - protein search, using sw model

Run on: November 14, 2003, 15:15:24 ; Search time 21 Seconds  
(without alignments)

2037,860 Million cell updates/sec

Title: US-10-339-656-2

Perfect score: 2410

Sequence: 1 MCIPLKASHSVSEEDTPSHY.....QRTIDEALQHPFDLKKK 445

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.76.\*  
2: PIR1.\*  
3: PIR2.\*  
4: PIR3.\*  
5: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1887	78.3	484	2	protein kinase CLK
2	1837	76.2	483	2	protein kinase STY
3	1727	71.7	484	2	protein kinase (BC
4	1417	58.8	499	2	protein kinase CLK
5	1321	54.8	490	2	protein kinase CLK
6	1317	54.6	490	2	protein kinase CLK
7	1222	50.7	517	2	protein kinase Dar
8	965.5	40.1	903	2	probable protein k
9	876	36.3	306	2	protein kinase STY
10	848.5	35.2	427	2	protein kinase APC
11	839	34.8	431	2	protein kinase PKI
12	830	34.4	467	2	protein kinase, 54
13	790	32.8	400	2	protein kinase AME
14	767	31.8	575	2	protein kinase hmo
15	767	31.8	690	2	probable protein k
16	718.5	29.8	737	1	probable serine/th
17	530	22.0	543	2	probable serine/th
18	524	21.7	1087	2	hypothetical prote
19	523	21.7	508	2	hypothetical prote
20	523	21.7	817	2	hypothetical prote
21	504.5	20.9	589	2	protein kinase DYR
22	478.5	19.9	629	2	protein kinase DYR
23	477	19.8	1157	2	hypothetical prote
24	474	19.7	948	2	hypothetical prote
25	472.5	19.6	754	2	Down-syndrome-crit
26	448.5	18.6	1189	2	homeodomain-intera
27	443.5	18.4	1155	2	probable protein k
28	441	18.3	1192	2	homeodomain-intera
29	437	18.1	642	2	protein kinase-lik

30	435	18.0	570	2	probable protein k
31	432	17.9	1191	2	serine/threonine p
32	431.5	17.9	807	2	protein kinase YAK
33	430	17.8	821	2	hypothetical prote
34	429	17.8	1457	2	protein kinase yak
35	419	17.4	1209	2	homeodomain-intera
36	406.5	16.9	477	2	mRNA splicing-asso
37	401	16.6	1093	2	protein B0464.5a l
38	398	16.5	775	2	hypothetical prote
39	392	16.3	1087	2	hypothetical prote
40	389	16.1	440	2	probable protein k
41	380.5	15.8	438	2	protein kinase hom
42	367.5	15.2	782	2	protein B0464.5c l
43	343	14.2	333	2	protein kinase (BC
44	338.5	14.0	523	2	serine protein kin
45	334	13.9	361	2	mitogen-activated

#### ALIGNMENTS

##### RESULT 1

S53641  
protein kinase CLK (BC 2.7.1.-) - human

C/Species: Homo sapiens (man)  
C/Date: 15-Jul-1995 #sequence #revision 01-Sep-1995 #text\_change 03-May-2002

C/Accession: S53641  
R/Hanes, J.; von der Kammer, H.; Klaudiny, J.; Scheit, K.H.

J. Mol. Biol. 244, 665-672, 1994

A/Title: Characterization by cDNA cloning of two new human protein kinases. Evidence b

A/Reference number: S53637; M01D:95082033; PMID:7990150

A/Accession: S53641  
A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Cross-references: GB:U92219; NID:9632963; PID:AA61480.1; PID:9632964

C/Superfamily: human protein kinase CLK1; protein kinase homology

C/Keywords: alternative splicing; phosphotransferase

F:159-430/Domain: protein kinase homology <kin>

Query Match 78.3%; Score 1887; DB 2; Length 484;

Best Local Similarity 82.1%; Pred. No. 5e-83;

Matches 352; Conservative 30; Mismatches 45; Indels 2; Gaps 2;

QY	18	SHYLSRLNERNDRYVDEYRNDYCEGYPRHYHNDISGRIRHSSKSSVSGSRSSP	77
DB	54	SHYLSRLNERNDRYVDEYRNDYCEGYPRHYHNDISGRIRHSSKSSVSGSRSSY	113
QY	78	KKK-RNRKCSH-QSRKSHRRKRSRSEDDEBGLICSGDVLARYEIVDTLGECAF	135
DB	114	KSKRIHSHRSHRSHRSHRSHRSHRSHRSHRSHRSHRSHRSHRSHRSHRSHRSHR	173
QY	136	KVVECIDHMDGMHVAVKIVXNNGRYRPARSEIQVLEHLSNSTDPSVFRVQMLEWFD	195
DB	174	KVVECIDHMDGMHVAVKIVXNNGRYRPARSEIQVLEHLSNSTDPSVFRVQMLEWFD	233
QY	196	HEHVCIVFELGLSYDFIKNSFLPQIDHIRQAAVQICISINFLHNKLTHTDLKEN	255
DB	234	HEHVCIVFELGLSYDFIKNSFLPQIDHIRQAAVQICISINFLHNKLTHTDLKEN	293
QY	256	ILFVSDYVVKYNSPKRDEKLTNTDIDKVDYFGSATDDEBHSITVSTRYRAPEVILA	315
DB	294	ILFVSDYVVKYNSPKRDEKLTNTDIDKVDYFGSATDDEBHSITVSTRYRAPEVILA	353
QY	316	LGMSQPCVWSIGCLILEYVGFVFPQHDSEKHLAMMERILGPIQGMIOKTRKKF	375
DB	354	LGMSQPCVWSIGCLILEYVGFVFPQHDSEKHLAMMERILGPIQGMIOKTRKKF	413
QY	376	HNOLMDHSSAGRYRRCKPLKEFMLCHDEBEKLPDLVRMLLEPTORTIDEALQ	435
DB	414	HNOLMDHSSAGRYRRCKPLKEFMLCHDEBEKLPDLVRMLLEPTORTIDEALQ	473
QY	436	HEFIDLK 444	

Db 474 HPFDLKK 482

# RESULT 2

A39676 protein kinase STY (EC 2.7.1.-) - mouse

N/Alternate names: protein kinase clk

C/Species: Mus musculus (house mouse)

C/Date: 08-Nov-1991 #sequence\_revision 08-Nov-1991 #text\_change 03-May-2002

C/Accession: A39676

R/Howell, B.W.; Afar, P.E.H.; Lew, J.; Douville, E.M.J.; Icelly, P.L.E.; Gray, D.A.; Bell, M.L. Cell Biol. 11, 568-572, 1991

A/Title: STY, a tyrosine-phosphorylating enzyme with sequence homology to serine/threonine

A/Reference number: A39676; PMID:1986248

A/Accession: A39676

A/Molecule type: mRNA

A/Residues: 1-483 <HOM>

A/Cross-references: GB:M38381, NID:g201070; P/DN:AAA0151.1; P/D:g201071

R/Ben-David, Y.; Letwin, K.; Tamock, L.; Bernstein, A.; Pawson, T.

EMBO J. 10, 317-325, 1991

A/Title: A mammalian protein kinase with potential for serine/threonine and tyrosine pho

A/Reference number: S13364; PMID:91122038; PMID:1825055

A/Accession: S13364

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-378, P, 380-483 <BEN>

A/Note: the sequence from Fig. 2 is inconsistent with that shown in Fig. 1 in having 448

C/Superfamily: human protein kinase clk; protein kinase homology

C/Keywords: phosphotransferase; serine/threonine-specific protein kinase

F:158-429/Domain: protein kinase homology <KIN>

Query Match 76.2%; Score 1837; DB 2; Length 483;

Best Local Similarity 79.7%; Pred. No. 1.2e-80;

Matches 349; Conservative 32; Mismatches 51; Indels 6; Gaps 4;

9 HVEEDTHSHYLSERDNDYRDYDEYNDYCEGYVPRHYRDIESGRIHCSGS 68

48 HSKTID--SYLSESSINIEKAYHSRRYDEYNDYM-GYEGHPGEGRSYOMHSSKS 103

69 SVSRSSRSPPKRR-RNR-HCSHQSRSKSHRRKRSRIEDDEBGLICQSGDVLARYEIV 126

104 SGRSGRSSYKSGRRRHHTSQHSHGKSHRRKRSVDEDEBGLICQSGDVLARYEIV 163

127 DTLSGAGRGKVECTIDHGMQMHVAVKIVKNGRYREARSEIOYLEHNLSTDPNSVPRC 186

164 DTLSGAGRGKVECTIDHGMQMHVAVKIVKNGRYREARSEIOYLEHNLSTDPNSVPRC 223

187 VQMLPMPHGHVCIIVFELIGSTYDFIKENSLFPQIDHITQMAVOICQINFLHANKL 246

224 VQMLPMPHGHVCIIVFELIGSTYDFIKENSLFPQIDHITQMAVOICQINFLHANKL 283

247 THTDLPENILFVKSDDYVVKYKSNKMRDERTLKNTDIKVDGSGATYDEHSTLVSTRH 306

284 THTDLPENILFVKSDDYVVKYKSNKMRDERTLKNTDIKVDGSGATYDEHSTLVSTRH 343

307 YRAPVILALGMSQCDVMSIGCIIIEYLGFTVQTHDSKHEHLAMERITLGPQHMIO 366

344 YRAPVILALGMSQCDVMSIGCIIIEYLGFTVQTHDSKHEHLAMERITLGPQHMIO 403

367 KTRRRKYFHHNQLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLFVLVRMLDYPTQ 426

404 KTRRRKYFHHNQLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLFVLVRMLDYPTQ 463

427 RITLDEALQHPFDLKK 444

464 RITLDEALQHPFDLKK 481

# RESULT 3

A38643 protein kinase (EC 2.7.1.37) cdc2/cdc28-like - human

C/Species: Homo sapiens (man)

C/Date: 21-Feb-1992 #sequence\_revision 21-Feb-1992 #text\_change 03-May-2002

C/Accession: A38643

R/Johnson, K.W.; Smith, K.A.

J. Biol. Chem. 266, 3407-3407, 1991

A/Title: Molecular cloning of a novel human cdc2/CDC28-like protein kinase.

A/Reference number: A38643; PMID:1704889

A/Accession: A38643

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-454 <JOH>

A/Cross-references: GB:M59287

C/Genetics:

A/Map position: 7q31-7q31

C/Superfamily: human protein kinase clk; protein kinase homology

C/Keywords: phosphotransferase

F:129-400/Domain: protein kinase homology <KIN>

Query Match 71.7%; Score 1727; DB 2; Length 454;

Best Local Similarity 76.6%; Pred. No. 1.9e-75;

Matches 327; Conservative 26; Mismatches 46; Indels 28; Gaps 3;

18 SHYLEARSINERDNDYRDYDEYNDYCEGYVPRHYRDIESGRIHCSGSVSRSSP 77

54 SHYLEARSINERDNDYRDYDEYNDYCEGYVPRHYRDIESGRIHCSGSVSRSSP 113

78 KKKRNRHCSHQSRSKSHRRKRSRIEDDEBGLICQSGDVLARYEIVDTLGEAGFGRV 137

114 KSKRHRIH-----HSTSHRSH-----GD-----EIVDTLGEAGFGRV 145

138 VECIDHGMQMHVAVKIVKNGRYREARSEIOYLEHNLSTDPNSVPRVQMLEWDHNG 197

146 VECIDHGMQMHVAVKIVKNGRYREARSEIOYLEHNLSTDPNSVPRVQMLEWDHNG 205

198 HVCIVFELIGSTYDFIKENSLFPQIDHITQMAVOICQINFLHANKLTHDLPENIL 257

206 HVCIVFELIGSTYDFIKENSLFPQIDHITQMAVOICQINFLHANKLTHDLPENIL 265

258 FYKSDYVVKYKSNKMRDERTLKNTDIKVDGSGATYDEHSTLVSTRHRAPEVILALG 317

266 FYKSDYVVKYKSNKMRDERTLKNTDIKVDGSGATYDEHSTLVSTRHRAPEVILALG 325

318 WSQPCDVMSIGCIIIEYLGFTVQTHDSKHEHLAMERITLGPQHMIOKTRRRKYFHHN 377

326 WSQPCDVMSIGCIIIEYLGFTVQTHDSKHEHLAMERITLGPQHMIOKTRRRKYFHHN 385

378 QLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLFVLVRMLDYPTQRTLDEALQHP 437

386 QLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLFVLVRMLDYPTQRTLDEALQHP 445

438 FFDLKK 444

446 FFDLKK 452

RESULT 4

A38637 protein kinase clk2, long splice form (EC 2.7.1.-) - human

C/Species: Homo sapiens (man)

C/Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 03-May-2002

C/Accession: S53637; T08825

R/Hanes, J.; von der Kammer, H.; Klautz, J.; Schelt, K.H.

J. Mol. Biol. 244, 665-672, 1994

A/Title: Characterization by cDNA cloning of two new human protein kinases. Evidence by

A/Reference number: S53637; PMID:95082033; PMID:7990150

A/Accession: S53637

A/Molecule type: mRNA

A/Residues: 1-499 <HAN>

A/Cross-references: GB:I29218; NID:g632967; P/DN:AAA61482.1; P/D:g632968

R/Minfeld, S.L.; Tayeh, N.; Martin, B.M.; Gims, E.I.; Sidransky, B.

Genome Res. 7, 1020-1026, 1997

A/Title: Identification of three additional genes contiguous to the glucocorticoid

A:Reference number: 216482; MUID:97474796; PMID:9331372  
 A:Accession: T08825  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-499 <MIN>  
 A:Cross-references: GB:A023268; NID:92564910; PIDN:AAC51817.1; PID:92564911  
 C:Comment: The short splice form of this protein (see PIR:S53638) lacks the protein kinase domain.  
 C:Gene: clk2  
 C:Function:  
 A:Description: Phosphorylates Ser, Thr, and Tyr residues on proteins in the spliceosomal C:Superfamily: human protein kinase clk1; protein kinase homology  
 C:Keywords: alternative splicing; phosphotransferase  
 F:161-440/Domain: protein kinase homology <KIN>

Query Match 58.8%; Score 1417; DB 2; Length 499;  
 Best Local Similarity 59.7%; Pred. No. 1.1e-60;  
 Matches 259; Conservative 69; Mismatches 98; Indels 8; Gaps 3;

```

18 SHYLEARS-----LNRDVRDRRYVDEY-RNDYCEGYVRHHRDIESGRIHCSKSVR 71
50 SHVRSRSSYDRSSDRRYDRYCGSYRNDYSRDGDAYDTDRHSTYEGRENSSYR 109
72 SRRSSPK-RKENRHCSSHQSRKSHRRKRSIENDEEGHLICQSGDYLARAYIVDTL 129
110 SQRSSRRKRRRRRRRSRTSRSSSQSSRAKSVEDAEGLIYHGDMLQERYELIVSTL 169
130 GEGARKEVVECTIDHGDGMHVAIKYKNGRYEARSETQVLEHLNSTDPNSVFCVOM 189
170 GGSTFGRRVQCVDRHRRGARVALKIKYNEKYEARLEINVLKINEXDPMKNCVQM 229
190 LEMFPHGHVCIVFELLGISTYDEIKENSFLPRQIDHITQMAVQIQSINFLHANKLTHT 249
230 FWMFDHGHMCISFELGISTFDELKNNYLPPIHQVRHMAFQICQAVKFLHNDLTHT 289
250 DLKPNILFVKSDYVYKNSKMKRDERLTAKNTDIAKVDGSAIVYDEHSTLYSTRHYA 309
290 DLKPNILFVNSDYELTYLTKKRDERSKSTAVRVVDGSAITFDEHSTLYSTRHYA 349
310 PEVITLALGWSQPCDWSIGICILIEYVIGFTVFQTHDSKHEMLAMERILGPIPOHMIOKTR 369
350 PEVITLALGWSQPCDWSIGICILIEYVIGFTVFQTHDNHRLAMERILGPIPSRMIRKTR 409
370 KRKYFHHNOLDWDEHSSAGRYRRCKPIKEFPLCHDEHEKLFDLVRMLEYDPTQRI 429
410 KQKYRGRGLDWEDENTSAGRYRENCKPLRRLYLSAEHQLFDLIESMLEYEPKRLT 469
430 IDEALQHPFDLLK 443
470 LGBALQHPFFARLR 483
  
```

RESULT 5  
 S53639  
 protein kinase clk3 (EC 2.7.1.-) - human  
 C:Species: Homo sapiens (man)  
 C:Date: 28-Oct-1996 #sequence\_revision 07-Feb-1997 #text\_change 03-May-2002  
 C:Accession: S53639; S71040  
 R:Hanes, J.; von der Kammer, H.; Klaudiny, J.; Scheit, K.H.  
 J. Mol. Biol. 244, 665-672, 1994  
 A:Title: Characterization by cDNA cloning of two new human protein kinases. Evidence by  
 A:Reference number: S53637; MUID:95082033; PMID:7990150  
 A:Accession: S53639  
 A:Molecule type: mRNA  
 A:Residues: 1-490 <HAN>  
 A:Cross-references: GB:L29217  
 R:Hanes, J.; von der Kammer, H.; Klaudiny, J.; Scheit, K.H.  
 submitted to the EMBL Data Library, January 1995  
 A:Reference number: S71040  
 A:Accession: S71040  
 A:Molecule type: mRNA  
 A:Residues: 1-131, 'TG', 134-490 <HAN>  
 A:Cross-references: EMBL:L29217; NID:9632971; PIDN:AAA61484.1; PID:9632972

C:Superfamily: human protein kinase clk1; protein kinase homology  
 C:Keywords: alternative splicing; phosphotransferase  
 F:154-433/Domain: protein kinase homology <KIN>

Query Match 54.8%; Score 1321; DB 2; Length 490;  
 Best Local Similarity 55.3%; Pred. No. 3.7e-56;  
 Matches 251; Conservative 60; Mismatches 101; Indels 42; Gaps 6;

```

4 PLEASHSVEEDTPSHYLEARSINERDYDRRYVDEYRNDYCEGYVP-----R 51
43 PPRRSRSRSHDRP-----YQRYRRRSDTYR---CEERSFPEDYGGPESR 90
52 HYHNDIEG-YRI-----HCKSSVRSRRSPKRNHCSHQSRKSHRRKRSRSTED 105
91 HRRRSRERGPYRTRKAHHC-----RTRSCSSASRSRQSSKSSRSRVED 138
106 DEEGHLICQSGDYLARAYIVDTLGEAGKRYVECTIDHGDGMHVAIKYKNGRYEAR 165
139 DKEGHLVCRIGDMLQSRYEIVGNLGGTGKVVCEIDHARGKQVALKITRNQKYEAA 198
166 RSEIQLVLEHINSTDPNSVFCVOMLEMPHGHVCIVFELLGISTYDFIKENSFLPRQID 225
199 RLEINVLKIKKEDKXNKFLCVLSDWTFPHGMCTAFELGKNTFEFLKENNFQYPLP 258
226 HIRQAAVQIQSINFLHANKLTHTDILKPNILFVKSDYVYKNSKMKRDERLTAKNTDIX 285
259 HVHMAVQICHALFELHNLTHTDILKPNILFVNSEFETLYNHSKCEKSVYNTSIRV 318
286 VQFSATYDDEHSTLYSTRHARPEVITLALGWSQPCDWSIGICILIEYVIGFTVFQTHD 345
319 ADGSAITFPEHHTTIVATRRHYRPEVITLALGWSQPCDWSIGICILIEYVIGFTVFQTHD 378
346 SKHEMLAMERILGPIPOHMIOKTRKRYFHHNOLDWDEHSSAGRYRRCKPIKEFPLCH 405
379 NREHLVMAEKILGPIPSHILHTRKOKIYKKGGLVWDEHSSDGRYVENCKPLKSYMLQD 438
406 DEEHEKLFDLVRMLEYDPTQRIIDEALQHPFF 439
439 SLEHVQLFDLMRMLFEPDQRIITLAEALHPFF 472
  
```

RESULT 6  
 S70352  
 protein kinase clk3 (EC 2.7.1.-), testis-specific - rat  
 N:Alternate names: LAMMER kinase  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 21-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 03-May-2002  
 C:Accession: S70352  
 R:Becker, W.; Kentrup, H.; Heukelbach, J.; Joost, H.G.  
 Biochim. Biophys. Acta 1312, 63-67, 1996  
 A:Title: cDNA cloning and characterization of rat CLK3, a LAMMER kinase predominately  
 A:Reference number: S70352; MUID:96271481; PMID:8679717  
 A:Accession: S70352  
 A:Molecule type: mRNA  
 A:Residues: 1-490 <BEC>  
 A:Cross-references: EMBL:X94351; NID:g1149536; PIDN:CAA64076.1; PID:g1149537  
 C:Gene: clk3  
 C:Superfamily: human protein kinase clk1; protein kinase homology  
 C:Keywords: phosphotransferase; protein kinase  
 F:154-433/Domain: protein kinase homology <KIN>

Query Match 54.6%; Score 1317; DB 2; Length 490;  
 Best Local Similarity 55.1%; Pred. No. 5.8e-56;  
 Matches 250; Conservative 61; Mismatches 101; Indels 42; Gaps 6;

```

4 PLEASHSVEEDTPSHYLEARSINERDYDRRYVDEYRNDYCEGYVP-----R 51
43 PPRRSRSRSHDRP-----YQRYRREHSDTYR---CEERSFPGEPCYSSSR 90
52 HYHNDIEG-YRI-----HCKSSVRSRRSPKRNHCSHQSRKSHRRKRSRSTED 105
91 HRRRSRERGPYRTRKAHHC-----RTRSCSSASRSRQSSKSSRSRVED 138
106 DEEGHLICQSGDYLARAYIVDTLGEAGKRYVECTIDHGDGMHVAIKYKNGRYEAR 165
139 DKEGHLVCRIGDMLQSRYEIVGNLGGTGKVVCEIDHARGKQVALKITRNQKYEAA 198
166 RSEIQLVLEHINSTDPNSVFCVOMLEMPHGHVCIVFELLGISTYDFIKENSFLPRQID 225
199 RLEINVLKIKKEDKXNKFLCVLSDWTFPHGMCTAFELGKNTFEFLKENNFQYPLP 258
226 HIRQAAVQIQSINFLHANKLTHTDILKPNILFVKSDYVYKNSKMKRDERLTAKNTDIX 285
259 HVHMAVQICHALFELHNLTHTDILKPNILFVNSEFETLYNHSKCEKSVYNTSIRV 318
286 VQFSATYDDEHSTLYSTRHARPEVITLALGWSQPCDWSIGICILIEYVIGFTVFQTHD 345
319 ADGSAITFPEHHTTIVATRRHYRPEVITLALGWSQPCDWSIGICILIEYVIGFTVFQTHD 378
346 SKHEMLAMERILGPIPOHMIOKTRKRYFHHNOLDWDEHSSAGRYRRCKPIKEFPLCH 405
379 NREHLVMAEKILGPIPSHILHTRKOKIYKKGGLVWDEHSSDGRYVENCKPLKSYMLQD 438
406 DEEHEKLFDLVRMLEYDPTQRIIDEALQHPFF 439
439 SLEHVQLFDLMRMLFEPDQRIITLAEALHPFF 472
  
```





protein kinase STY (EC 2.7.1.-) [imported] - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Jun-2000  
 C:Accession: I49068  
 R:Duncan, P.I.; Howell, B.W.; Marius, R.M.; Drmanic, S.; Douville, E.M.; Bell, J.C.  
 J. Biol. Chem. 270, 21524-21531, 1995  
 A:Title: Alternative splicing of STY, a nuclear dual specificity kinase.  
 A:Reference number: I49068; MUID:95394904; PMID:7665564  
 A:Accession: I49068  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-306 <RES>  
 A:Cross-references: EMBL:U11054; NID:9507917; PID:AAA82185.1; PID:9507918  
 C:Genetics:  
 A:Introns: 129/3; 160/1; 221/2  
 C:Keywords: phosphotransferase

Query Match 36.3%; Score 876; DB 2; Length 306;  
 Best Local Similarity 74.2%; Pred. No. 3.2e-35;  
 Matches 173; Conservative 21; Mismatches 33; Indels 6; Gaps 4;  
 QY 9 HSEVEDTHPSHYLEARSINERDYDRRYVDEYNDYCEGYVPHYHNDIESGYRIHCKSKS 68  
 D 48 HSKTTD---SYLESRSINERKAYHRRYVDEYNDY-GEYEPHGYEPGSRQOMSSKKS 103  
 QY 69 SVNRSSSPKPK-RNR-HGSHQGRSKSRFRKRSRSTEDDEBGLTCQSGDVLRARYEIV 126  
 D 104 SGRGRSSYSKSRHSHHTSOHSHGSKSRKRSRYVDEDEBGLTCQSGDVLSARYEIV 163  
 QY 127 DTLEGAFAKVECTIDHGMDFMVAIVKVGGRFARSAEIOVLEHINSTDPNSVFR 186  
 D 164 DTLEGAFAKVECTIDHGMDFMVAIVKVGGRFARSAEIOVLEHINSTDPNSVFR 223  
 QY 187 VQMLEFDFHGHVCIIVELLGLSTYDFIKENSFLPFOIDIRQMAVOICQSN 239  
 D 224 VQMLEFDFHGHVCIIVELLGLSTYDFIKENSFLPFOIDIRQMAVOICQSN 276

RESULT 10  
 T05560  
 N:protein kinase Afc2 (EC 2.7.1.-) - Arabidopsis thaliana  
 A:Alternate names: protein P22K18.60  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 17-Mar-2000  
 C:Accession: T05560  
 R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohnsels, J.; Mewes, H.W.; Mayer, K.F.  
 submitted to the Protein Sequence Database, February 1999  
 A:Reference number: Z15419  
 A:Accession: T05560  
 A:Molecule type: DNA  
 A:Residues: 1-427 <BRV>  
 A:Cross-references: EMBL:AJ035356  
 A:Experimental source: cultivar Columbia, BAC clone P22K18  
 C:Genetics:  
 A:Gene: Afc2  
 A:Map position: 4  
 A:Introns: 34/3; 98/1; 107/1; 158/2; 174/3; 213/1; 243/3; 294/1; 315/3; 348/2; 382/3  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  
 C:Keywords: ATP; phosphotransferase; protein kinase

Query Match 35.2%; Score 848.5; DB 2; Length 427;  
 Best Local Similarity 48.1%; Pred. No. 8.8e-34;  
 Matches 167; Conservative 70; Mismatches 97; Indels 13; Gaps 8;  
 QY 104 EDEBEGHLCQSGVLRARYEIVDTLEGAFAKVECTIDHGMDFMVAIVKVGGRFAR 163  
 D 79 EDDDDGYIFELGDDLPFRYKYSKMGEGFGVLECDREKEM-VAVYIVGVYKRYR 137  
 QY 164 AARSEIOVLEHINSTDPNSVFRQVMLEFDFHGHVCIIVELLGLSTYDFIKENSFLP 223  
 D 138 AAMEIEMLOQLGKHDSG-NRCVQINWFDYRNHICITVEFKGSSILDLRKNVRSFP 196

QY 224 IDHIRQMAVOICQSNIFLHNNKLTHTDLKPENILFVKSQDVY-VKY-NSKKMRD---ERT 277  
 D 197 IDLVAREIQGLCEVAFMHDLRMHTDLKPENILLVSSDYKIPYKSSRLQRFQVCYRV 256  
 QY 278 LKNTDIAKVVDFSSATYDDEHNSSTVTHHAPRYTALGMSQCPDWSSICILIEYVLG 337  
 D 257 PMSAIKXIDFSTYIERQDDQYIYSTHYAPVILGLGMSPCDWSSVCCIIYELCTG 316  
 QY 338 FTVPQTHSKELAMERILIGIPQHMIOKT--RRKYFHHNQLDWDHSSAGRYVRRRC 395  
 D 317 EALFQTHENLEHAMERVLPFPQOMKKYDRHSEKTVRRGRDLDPGQATRSLSKAVL 376  
 QY 396 K--PLKEFMLC-DEHEKLFEDLYRMDEYDPTORTLDEALQHPFF 439  
 D 377 KLPRLQNLIMQVDSAGELINMVGRLRFPDSERITAREALRHPFF 423

RESULT 11  
 T04125  
 N:protein kinase PK12 (EC 2.7.1.-), ethylene-induced - common tobacco  
 C:Species: Nicotiana tabacum (common tobacco)  
 C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
 C:Accession: T04125  
 R:Jessa, G.; Raz, V.; Savaldi, S.; Fluhr, R.  
 Plant Cell 8, 2223-2234, 1996  
 A:Title: PK12, a plant dual-specificity protein kinase of the LAMMER family, is regulated by ethylene.  
 A:Reference number: Z15228; MUID:97143872; PMID:8989879  
 A:Accession: T04125  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-431 <SES>  
 A:Cross-references: EMBL:U73937; NID:92911279; PID:AA04324.1; PID:92911280  
 A:Experimental source: cultivar Samsun NN  
 C:Genetics:  
 A:Gene: PK12  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  
 C:Keywords: ATP; phosphotransferase; protein kinase  
 P:94-381/Domain: protein kinase homology <KIN>

Query Match 34.8%; Score 839; DB 2; Length 431;  
 Best Local Similarity 39.9%; Pred. No. 2.5e-33;  
 Matches 175; Conservative 80; Mismatches 130; Indels 54; Gaps 10;  
 QY 29 RPYRRRYVDEYRANVYCGYVPRHNRHIDEGYRIHCKSKSVRSR----- 73  
 D 11 RPYMRPRKRRPRLDWDSHTPK-AQSGIYYQEGVSSSYVHSRLPDDHSLVYKGLAQ 69  
 QY 74 RSPKRRKRRHCSHQSSKSKSRKRSRIEDDEBGLTCQSGDVLRARYEIVDTLEGA 133  
 D 70 KSPPRR-----DDDDGHIMFELGENTLTRYIKLILKIGSGT 106  
 QY 134 PGVVECTIDHGMDFMVAIVKVGGRFARSAEIOVLEHINSTDPNSVFRQVMLEF 193  
 D 107 PGQVLECDREKEM-VAVYIVGVYKRYR-RCVOLRNMF 164  
 QY 194 DHGHVCIIVELLGLSTYDFIKENSFLPFOIDIRQMAVOICQSNIFLHNNKLTHTDLKP 253  
 D 165 DTRNHICLVFEKLGSLDPLFKNSYRAPVDVLAIEGRQLBCEAFHMDKLLHTDLKP 224  
 QY 254 ENILFVKSQDVY---VKTNSKKMRD---ERTLKNITDIKYVDFGSAFYDDEHNSSTVSTR 305  
 D 225 ENILFVSADYIKVPRYKTPWSHRSFSPKRLPKSAIKVLDIFGSTAYERPDDHNYVSTR 284  
 QY 306 HYRAPEVILALGMSQCPDWSSICILIEYVLGFTYFQTHDSKELAMERILIGIPQ 365  
 D 285 HYRAPEVILALGMSQCPDWSSICILIEYVLGFTYFQTHDSKELAMERILIGIPQ 344  
 QY 366 QKTRK--RKYFHHNQLDWDHSSAGRYVRRCK--PLKEFMLC-DEHEKLFEDLYRM 420  
 D 345 KVDVHAERVYARRGLDWPREGATRESISKVWKLPRLQNLVWQYVDHAGLIDLDLQGL 404  
 QY 421 EYDPTORTLDEALQHPFF 439

Db 405 RFPDPSIRMTAHDAALRHPFF 423

# RESULT 12

S71169

protein kinase, 54K (EC 2.7.1.-) - Arabidopsis thaliana

N:Alternate names: protein F4P12.270

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 03-May-2002

C:Accession: S71169, T45897

R:Kurumori, T.; Yamamoto, M.

submitted to the EMBL Data Library, January 1995

A:Description: A.thaliana genes encoding protein kinases of a new family.

A:Reference number: S71169

A:Accession: S71169

A:Molecule type: mRNA

A:Residues: 1-467 <KUR>

A:Cross-references: EMBL:D45354; NID:g642131; PID:g642132

R:Blocker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Queciet, F.; Salanoubat, M.

submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23016

A:Accession: T45897

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-467 <BLO>

A:Cross-references: EMBL:AL132966

A:Experimental source: cultivar Columbia; BAC clone F4P12

C:Genetics:

A:Map position: 3

A:Introns: 56/3; 115/1; 124/1; 175/2; 191/3; 230/1; 260/3; 313/1; 334/3; 367/2; 402/3

C:Superfamily: human protein kinase c1k1; protein kinase homology

C:Keywords: ATP; phosphotransferase; protein kinase

F:113-401/Domain: protein kinase homology <KIN>

Query Match 34.4%; Score 830; DB 2; Length 467;

Best Local Similarity 45.8%; Pred. No. 7.2e-33;

Matches 160; Conservative 71; Mismatches 102; Indels 16; Gaps 8;

105 DDEGHLICQSGDVLARAYEIVDTLGEAGFVCEIDHGMGMVAIVKIVNGVYREA 164

Db 97 DDKDHVYFVVDITLPRYQILSKMGEGTFGVLCEFDNKNKV-VAIIVISINKYREA 155

QY 165 ARSEIQVLEHLNSTDPNSVFCVOMLEMPDHGHVCIYFELLGLSTYDFIKNSFLPCOI 224

Db 156 AMIEIDVILQRLTRHVVGG-SRCVOIRMPEDYRNHICIVFEKLGPSLDFLRKNSVRSFPI 214

QY 225 DHIRMAVOICQSIINFLHNKLTHTDLKPENLIFVKSQDV---VKYNSKMRDEFTLKN 280

Db 215 DLVREIGRLQLESAVMHDLRLHTDLKPENLILVSSEYIKLPDYKFLSRPTKDSYFKN 274

QY 281 ----TDIKVDFGSATYDDEHSTLVSTRHAPAEVITLALGMSOPCDVWSIGCILLEYVL 336

Db 275 LPKSAIKLIDFGSTTFEHDQNIYVSTRHAPAEVITLALGMSOPCDVWSIGCILLEYVL 334

QY 337 GTTVQTHDSKHEHLAMERITLPIQHMIOKT-RKRYFHNN-QLDWDHSSAGRYVR 393

Db 335 GEALQTHENHDLAMERIVLGLPLPHVTLRADRSSEKFRGAKLDMPEGATSDSLKA 394

QY 394 --RCKPLKEFMLCH-DEEHEKLFDLVRMLDYDPORITLIDALOHPPF 439

Db 395 VMLKPLRLNLMQVHDHSAAGLLIDLGULRIYDIFERKAREALNHPFF 443

RESULT 13

T04460

protein kinase AMB3 (EC 2.7.1.-) - Arabidopsis thaliana

N:Alternate names: protein F4D11.140

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 17-Mar-2000

C:Accession: T04460

R:Bevan, M.; Benes, V.; Rechmann, S.; Borikova, D.; Ansoorge, W.; Hohelsel, J.; Mewes, H.W.

submitted to the Protein Sequence Database, April 1998

Query Match 31.8%; Score 767; DB 2; Length 575;

Best Local Similarity 41.5%; Pred. No. 8.5e-30;

Matches 141; Conservative 78; Mismatches 115; Indels 6; Gaps 3;

105 DDEGHLICQSGDVLARAYEIVDTLGEAGFVCEIDHGMGMVAIVKIVNGVYREA 164

Db 57 ISSGRIHCSKSVSR-----RSSPKRR--NRCSHQSKSHRRKRSIED 105

QY 2 IANGFE-SMDKRRVKKRPMTWDEAPAEPAKRAVYKGGSDGRILSPPLR-----DD 53

Db 106 DDEGHLICQSGDVLARAYEIVDTLGEAGFVCEIDHGMGMVAIVKIVNGVYREA 165

QY 54 DRDGHVFSIRDLNLTLPYKILSKMGEGTFGVLCEFDNKNKV-EYAIKILIRSIKKYRDA 112

Db 166 RSEIQVLEHLNSTDPNSVFCVOMLEMPDHGHVCIYFELLGLSTYDFIKNSFLPCOI 225

QY 113 MEIDVLQKLVKSDGRT-RCVOMKNWFYRNHICIVFEKLGPSLDFLRKNSYAFPLA 171

Db 226 HIRMAVOICQSIINFLHNKLTHTDLKPENLIFVKSQDV---VKYNSKMRDEFTLKN 280

QY 112 LVRDGCGQLLESAVMHDLRLHTDLKPENLILVSSEYIKLPDYKFLSRPTKDSYFKN 274

Db 281 TDIKVDFGSATYDDEHSTLVSTRHAPAEVITLALGMSOPCDVWSIGCILLEYVL 336

QY 222 SAIKLIDFGSTTFEHDQNIYVSTRHAPAEVITLALGMSOPCDVWSIGCILLEYVL 334

Db 341 FQTHDSKHEHLAMERITLPIQHMIOKT-RKRYFHNN-QLDWDHSSAGRYVR 393

QY 222 FQTHDSKHEHLAMERITLPIQHMIOKT-RKRYFHNN-QLDWDHSSAGRYVR 393

Db 335 GEALQTHENHDLAMERIVLGLPLPHVTLRADRSSEKFRGAKLDMPEGATSDSLKA 394

A:Reference number: Z15360

A:Accession: T04460

A:Molecule type: DNA

A:Residues: 1-400 <BEV>

A:Cross-references: EMBL:AL022537

A:Experimental source: cultivar Columbia; BAC clone F4D11

C:Genetics:

A:Gene: AMB3

A:Map position: 4

A:Introns: 30/3; 71/1; 80/1; 131/2; 147/3; 186/1; 216/3; 266/1; 287/3; 320/2; 355/3

A:Note: F4D11.140

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homo

C:Keywords: ATP; phosphotransferase; protein kinase

Query Match 32.8%; Score 790; DB 2; Length 400;

Best Local Similarity 41.7%; Pred. No. 4.9e-31;

Matches 169; Conservative 70; Mismatches 134; Indels 32; Gaps 11;

57 ISSGRIHCSKSVSR-----RSSPKRR--NRCSHQSKSHRRKRSIED 105

Db 2 IANGFE-SMDKRRVKKRPMTWDEAPAEPAKRAVYKGGSDGRILSPPLR-----DD 53

QY 106 DDEGHLICQSGDVLARAYEIVDTLGEAGFVCEIDHGMGMVAIVKIVNGVYREA 165

Db 54 DRDGHVFSIRDLNLTLPYKILSKMGEGTFGVLCEFDNKNKV-EYAIKILIRSIKKYRDA 112

QY 166 RSEIQVLEHLNSTDPNSVFCVOMLEMPDHGHVCIYFELLGLSTYDFIKNSFLPCOI 225

Db 113 MEIDVLQKLVKSDGRT-RCVOMKNWFYRNHICIVFEKLGPSLDFLRKNSYAFPLA 171

QY 226 HIRMAVOICQSIINFLHNKLTHTDLKPENLIFVKSQDV---VKYNSKMRDEFTLKN 280

Db 112 LVRDGCGQLLESAVMHDLRLHTDLKPENLILVSSEYIKLPDYKFLSRPTKDSYFKN 274

QY 281 TDIKVDFGSATYDDEHSTLVSTRHAPAEVITLALGMSOPCDVWSIGCILLEYVL 336

Db 222 SAIKLIDFGSTTFEHDQNIYVSTRHAPAEVITLALGMSOPCDVWSIGCILLEYVL 334

QY 341 FQTHDSKHEHLAMERITLPIQHMIOKT-RKRYFHNN-QLDWDHSSAGRYVR 393

Db 222 FQTHDSKHEHLAMERITLPIQHMIOKT-RKRYFHNN-QLDWDHSSAGRYVR 393

QY 336 KPLKEFMLCH-DEEHEKLFDLVRMLDYDPORITLIDALOHPPF 439

Db 352 DRLKMGSKRNDVNTSRFADLLYGLAYDPSERTIANEALHHPFF 396

RESULT 14

JC7794

lammer kinase homolog protein 1, Lkh1 protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 02-Apr-2002 #sequence\_revision 02-Apr-2002 #text\_change 02-Apr-2002

C:Accession: JC7794

R:Kim, K.H.; Cho, Y.M.; Kang, W.H.; Kim, J.H.; Byun, K.H.; Park, Y.D.; Bae, K.S.; Park,

Biochem. Biophys. Res. Commun. 289, 1237-1242, 2001

A:Title: Negative regulation of filamentous growth and flocculation by Lkh1, a fission

A:Reference number: JC7794

A:Accession: JC7794

A:Molecule type: mRNA

A:Residues: 1-575 <KIM>

A:Cross-references: GB:AF334941

A:Experimental source: strain SP286, ED665

C:Comment: This protein, a homolog of the dual-specificity protein kinase of the LAMMER

ast, by repressing expression of cell surface proteins for flocculation and morphogenesis

C:Genetics:

A:Gene: Lkh1+

Query Match 31.8%; Score 767; DB 2; Length 575;

Best Local Similarity 41.5%; Pred. No. 8.5e-30;

Matches 141; Conservative 78; Mismatches 115; Indels 6; Gaps 3;

105 DDEGHLICQSGDVLARAYEIVDTLGEAGFVCEIDHGMGMVAIVKIVNGVYREA 164

Db 57 ISSGRIHCSKSVSR-----RSSPKRR--NRCSHQSKSHRRKRSIED 105

QY 2 IANGFE-SMDKRRVKKRPMTWDEAPAEPAKRAVYKGGSDGRILSPPLR-----DD 53

Db 106 DDEGHLICQSGDVLARAYEIVDTLGEAGFVCEIDHGMGMVAIVKIVNGVYREA 165

QY 54 DRDGHVFSIRDLNLTLPYKILSKMGEGTFGVLCEFDNKNKV-EYAIKILIRSIKKYRDA 112

Db 166 RSEIQVLEHLNSTDPNSVFCVOMLEMPDHGHVCIYFELLGLSTYDFIKNSFLPCOI 225

QY 113 MEIDVLQKLVKSDGRT-RCVOMKNWFYRNHICIVFEKLGPSLDFLRKNSYAFPLA 171

Db 226 HIRMAVOICQSIINFLHNKLTHTDLKPENLIFVKSQDV---VKYNSKMRDEFTLKN 280

QY 112 LVRDGCGQLLESAVMHDLRLHTDLKPENLILVSSEYIKLPDYKFLSRPTKDSYFKN 274

Db 281 TDIKVDFGSATYDDEHSTLVSTRHAPAEVITLALGMSOPCDVWSIGCILLEYVL 336

QY 222 SAIKLIDFGSTTFEHDQNIYVSTRHAPAEVITLALGMSOPCDVWSIGCILLEYVL 334

Db 341 FQTHDSKHEHLAMERITLPIQHMIOKT-RKRYFHNN-QLDWDHSSAGRYVR 393

QY 222 FQTHDSKHEHLAMERITLPIQHMIOKT-RKRYFHNN-QLDWDHSSAGRYVR 393

Db 335 GEALQTHENHDLAMERIVLGLPLPHVTLRADRSSEKFRGAKLDMPEGATSDSLKA 394

QY 394 --RCKPLKEFMLCH-DEEHEKLFDLVRMLDYDPORITLIDALOHPPF 439

Db 395 VMLKPLRLNLMQVHDHSAAGLLIDLGULRIYDIFERKAREALNHPFF 443

Job time : 22 secs

Db 222 DDDGHHKVVYNSKFMNRITVVRLLGHGFFGKVIQCYQOS--TGRGCAIKYTRALPKYREA 287

QY 165 ARSEIQVLEHLNSTDPNSVFRCYOMLEMHEDHGHGCIYELLGLSTYPIENSFLEPOI 224

Db 288 SLEELRFLQTLASHSDPNEKCKIQLRDYDYRKHLCIYTDLFGMSVGFELKNYYIPPL 347

QY 225 DHIROMAYQICOSINFLHNHKLHTYDLKPENLFPYKSDVYKYSKMRDERTLKNTDIK 28

Db 348 KHIOMLSQOLFKAFFVAFLHSGLVHTDLKPENLVLNSASRTIRLEPRNYSQVLSNCSIR 407

QY 285 VWDPGASTYDDEHSLTSTVTRHARAEVLALGMSQOPDWSIGCILLLEYLGFVTCOT 344

Db 408 LIDGSAITFEDEHSSVSTSTRYARAEIILGLGMSYPCQVWSIGCILLVELSTGALFOYH 467

QY 345 DSKENHLAMERILGPRIQHM--QKTRKKKFFHNQOLDMEHSSAGHYVR--RCKPLX 399

Db 468 EDSEHSLCQMEKIKIGPEDRNWISRSRSTQSRFKSGDKVRYPLSTNPKKSIYVLSQLTLE 527

QY 400 EFMILCHDEHEKFLFDVRYMLEDPQORTLDEALOHPEF 439

Db 528 QIRAVSSPEVALLDLKLKFPYVDPRKRIITAKAELAHPEF 567

## RESULT 15

Probable protein kinase - fission yeast (*Schizosaccharomyces pombe*)  
 C:Species: *Schizosaccharomyces pombe*  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Aug-2000  
 C:Accession: T38052  
 R:Lye, G.; Churchell, C.M.; Barrel, B.G.; Rajandream, M.A.; Walsh, S.V.  
 Submitted to the EMBL Data Library, February 1995  
 A:Reference number: 221765  
 A:Accession: T38052  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Cross-references: 1-690 <LIE>  
 A:Cross-references: EMBL:Z69239, PIDN:CAA93220.1, GSPDB:GM00066, SPDB:SPAC1D4.11d  
 A:Experimental source: strain 972h-; cosmid CID4  
 C:Genetics:  
 A:Gene: SPDB:SPAC1D4.11c  
 A:Map position: 1  
 C:Superfamily: yeast probable protein kinase KNS1; protein kinase homology

Query Match	31.8%;	Score 767;	DB 2;	Length 690;
Best Local Similarity	41.5%;	Pred. No. 1e-29;		
Matches 141;	Conservative 78;	Mismatches 115;	Indels 6;	Gaps 3;

```

OY      105  DDEGHILCOSGUVLAFARAEIVDLTGEGAFKVCVCIHDMDGMEVAKIYXNNGRREA 164
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      344  DDDGHYKVVNPKFNRYKTYVRLVRLSHGTFGKVICIGYQDS--TGRHCAIKVTRALPKREA 402

OY      165  ARSEIOVLHNLSTDNVSVFRCYOMIEMVDHGHVCYIVELIGLSTYDFIKENSFLPQI 224
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      403  SLIELRLQIASHDPTENKCIOTLRDYDYKHKICITDLTGMSVBPFLKNNYIPEPL 462

OY      225  DHIROMAYOICOSINFELHNKLTHTDLPKENLTFKSYDVVVKYNSKMRDERTLKNQDIK 284
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      463  KHIQMSQDLFKSVARLHSLGLVHDLKRENVLLVSNRSRTLRPLRYNSQKVLNSECIR 522

OY      285  VNDFGSATYDDDEHNSLTNSTRIHYRAEVLIAIGNSQPCDVMISGICILIEYLGETVPQTH 344
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      523  LIDFGSATPEDEHSSVNSTRIHYRAEIIILIGMSVPCDVMISGICILIELTQGLRQTH 582

OY      345  DSKEHLAMERLIGLPIPOHMI--OKTRKRYKFHHNOLDMDHESHSAGRYVR--RRCPLK 399
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      583  EDSBHLCMWEKILGPPDRAMIMSSRSTQSRPFKSDOKVRKYPPLSNTPKKSINYILQSLTLE 642

OY      400  EPMICHDEBEHKLPDLVRMLLEVDPPQRTITLDEALQHPPF 439
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      643  QIPAVSSPEVALLDLTKKTFVYUDDPKRRITLAEALMHPPE 682

```

Search completed: November 14, 2003, 15:19:14



Query Match 93.8%; Score 2261.5; DB 1; Length 481;  
 Best Local Similarity 93.8%; Pred. No. 5.7e-137;  
 Matches 420; Conservative 8; Mismatches 10; Indels 11; Gaps 2;

QY 8 SH-SVEEDTH-----PSHYLARSLSNERDYPDRRYVDEYRNDYCEGYVRHHRD 56  
 DB 33 SHSSTGNRHCQPHHQKDSCHLHARCLNERDYPDRRYVDEYRNDYCEGYVRHHRD 92  
 QY 57 IESGYRHCCKSSYRSRRSSPKRRKRNPHCSHQSRSKSHRRKRSIJEDEBGLICQSG 116  
 DB 93 VESTYRHCCKSSYRSRRSSPKRRKRNPCASHQSHSKSHRRKRSIJEDEBGLICQSG 152  
 QY 117 DVYRAREIVDTLGEAGFGVVECIDHGMCMHAYVTVKVGGRYRAASEIOVLEHLN 176  
 DB 153 DVYRAREIVDTLGEAGFGVVECIDHGMCMHAYVTVKVGGRYRAASEIOVLEHLN 212  
 QY 177 STDPSNVSRCVOMLEWMDHGHVCIYFELLGLSTYDPIKENSFLPFOIDHROMAYOICQ 236  
 DB 213 STDPSNVSRCVOMLEWMDHGHVCIYFELLGLSTYDPIKENSFLPFOIDHROMAYOICQ 272  
 QY 237 SINFLEHNKLTHTDLPKRNILFYKSDYVVKYNSKMRKDERTLKNTDIXVDFGSATYDDE 296  
 DB 273 SINFLEHNKLTHTDLPKRNILFYKSDYVVKYNSKMRKDERTLKNTDIXVDFGSATYDDE 332  
 QY 297 HNSTLVSTRHYRAPEVIALGMSGPCDVMISGICILEYVGTAFVPOHDSKELHAMMERI 356  
 DB 333 HNSTLVSTRHYRAPEVIALGMSGPCDVMISGICILEYVGTAFVPOHDSKELHAMMERI 392  
 QY 357 LSGIPQHMIOKTRKRYFHHNOLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 416  
 DB 393 LSGIPQHMIOKTRKRYFHHNOLDWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLV 452  
 QY 417 RRMLEVDPTORITLDEALQHPFDLKKK 445  
 DB 453 RRMLEVDPTORITLDEALQHPFDLKKK 481

RESULT 2  
 CLKT\_HUMAN STANDARD; PRT; 484 AA.  
 AC P49759;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protein kinase CLKT (EC 2.7.1.1-) (CLKT).  
 GN CLKT OR CLK.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91139618; PubMed=1704889;  
 RA Johnson K.W., Smith K.A.;  
 RT "Molecular cloning of a novel human cdc2/CD28-like protein kinase.";  
 RT J. Biol. Chem. 266:3402-3407(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95082033; PubMed=7990150;  
 RA Hanes J.J., der Kammer H., Klaudiny J.J., Schelt K.H.;  
 RT "Characterization by cDNA cloning of two new human protein kinases.  
 RT Evidence by sequence comparison of a new family of mammalian protein  
 RT kinases.";  
 RT J. Mol. Biol. 244:665-672(1994).  
 CC -!- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS  
 CC OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF  
 CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA  
 CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS: Nucleic.  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;

CC IsoId=P49759-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P49759-2; Sequence=VSP\_004852, VSP\_004853;  
 CC Note=Lacks the kinase domain;  
 CC -!- PIM: Autophosphorylates on all three types of residue (By  
 CC similarity).  
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; L29219; AAA61480.1; -;  
 CC EMBL; L29222; AAB59459.1; -;  
 CC PIR; S53641; S53641.  
 CC HSSP; Q00534; 1B18.  
 CC Genew; HGNC:2068; CLKT.  
 CC GK; P49759; -;  
 CC MIM; 601951; -;  
 CC GO; GO:0004715; F:non-membrane spanning protein tyrosine kina. . .; TAS.  
 CC GO; GO:0004674; F:protein serine/threonine kinase activity; TAS.  
 CC GO; GO:0008283; P:cell proliferation; TAS.  
 CC GO; GO:000074; P:regulation of cell cycle; TAS.  
 CC InterPro; IPR00719; Prot\_kinase.  
 CC InterPro; IPR002290; Ser\_thr\_kinase.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 1.  
 CC SMART; SM00220; S\_TKc; 1.  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 CC PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC Transferrase; Serine/threonine-protein kinase; ATP-binding;  
 CC Tyrosine-protein kinase; Phosphorylation; Nuclear protein;  
 CC KW alternative splicing.  
 CC FT DOMAIN 161 477 PROTEIN KINASE.  
 CC FT NP\_BIND 167 175 ATP (BY SIMILARITY).  
 CC FT BINDING 191 191 ATP (BY SIMILARITY).  
 CC FT ACT\_SITE 288 288 BY SIMILARITY.  
 CC FT VARSPPLIC 131 136 KSHRK -> MKLIL (in isoform Short).  
 CC FT FTID=VSP\_004852.  
 CC FT VARSPPLIC 137 484 Missing (in isoform Short).  
 CC FT FTID=VSP\_004853.  
 CC FT SEQUENCE 484 AA; 57205 MW; 304B584486AD0A6B CRC64;

Query Match 78.3%; Score 1887; DB 1; Length 484;  
 Best Local Similarity 82.1%; Pred. No. 3.7e-113;  
 Matches 352; Conservative 30; Mismatches 45; Indels 2; Gaps 2;

QY 18 SHYLSRSLSNERDYPDRRYVDEYRNDYCEGYVRHHRDIESGYRHCCKSSYRSRRSSP 77  
 DB 54 SHYLSRSLSNERDYPDRRYVDEYRNDYCEGYVRHHRDIESGYRHCCKSSYRSRRSSP 113  
 QY 78 KKK-RNRHCSSH-QSRSKSHRRKRSIJEDEBGLICQSGVLLARRYIVDTLGEAGRG 135  
 DB 114 KSKRIHSHSTRHSHGSHRRKRTSVDEDEBGLICQSGVLLARRYIVDTLGEAGRG 173  
 QY 136 KVVCEIDHGMCMHAYVTVKVGGRYRAASEIOVLEHLNSTDPSNVSRCVOMLEWMPH 195  
 DB 174 KVVCEIDHGMCMHAYVTVKVGGRYRAASEIOVLEHLNSTDPSNVSRCVOMLEWMPH 233  
 QY 196 HGHVCIYFELLGLSTYDPIKENSFLPFOIDHROMAYOICQINFLEHNKLTHTDLPKPN 255  
 DB 234 HGHVCIYFELLGLSTYDPIKENSFLPFOIDHROMAYOICQINFLEHNKLTHTDLPKPN 293  
 QY 256 ILFVSDYVVKYNSKMRKDERTLKNTDIXVDFGSATYDDEHSTLVSTRHYRAPEVIALA 315  
 DB 294 ILFVSDYVVKYNSKMRKDERTLKNTDIXVDFGSATYDDEHSTLVSTRHYRAPEVIALA 353



QY 316 LQMSQPCDVWISICILIEVYIGFTVOTHSKEHLAMMERILGPIQCHMIQTRKRYFH 375  
 DB 354 LQMSQPCDVWISICILIEVYIGFTVOTHSKEHLAMMERILGPIQCHMIQTRKRYFH 413  
 QY 376 HNOUWDEHSSAGRYVRRCKPLKEPMLCHDEHKLFDLYRMLERYPTOTRITLDEALQ 435  
 DB 414 HNRUWDEHSSAGRYVRRCKPLKEPMLCHDEHKLFDLYRMLERYPTOTRITLDEALQ 473  
 QY 436 HPEFDLLKK 444  
 DB 474 HPEFDLLKK 482

RESULT 3  
 CLK1 MOUSE STANDARD; PRT: 483 AA.

ID CLK1 MOUSE  
 AC P22518;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protein kinase CLK1 (EC 2.7.1.1-) (CLK) (Protein kinase STY).  
 OS CLK OR CLK OR STY.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OK NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91122038; PubMed=1825055;  
 RA Ben-David Y., Letwin K., Tannock L., Bernstein A., Pawson T.;  
 RT "A mammalian protein kinase with potential for serine/threonine and  
 RL tyrosine phosphorylation is related to cell cycle regulators.";  
 RL EMBO J. 10:317-325 (1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91094875; PubMed=1986248;  
 RA Howell B.W., Afar D.B., Lew J., Douville E.M., Icely P.L.,  
 RT Gray D.A., Bell J.C.;  
 RL "STY, a tyrosine-phosphorylating enzyme with sequence homology to  
 RT serine/threonine kinases.";  
 RL Mol. Cell. Biol. 11:568-572 (1991).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RX MEDLINE=95394904; PubMed=766564;  
 RA Duncan P.T., Howell B.W., Maris R.M., Drmanic S., Douville E.M.,  
 RT Bell J.C.;  
 RL "Alternative splicing of STY, a nuclear dual specificity kinase.";  
 RL J. Biol. Chem. 270:21524-21531 (1995).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RX MEDLINE=95082033; PubMed=7990150;  
 RA Hanes J.J., der Kammer H., Klaudiny J.J., Scheit K.H.;  
 RT Characterization by cDNA cloning of two new human protein kinases.  
 RT Evidence by sequence comparison of a new family of mammalian protein  
 RL kinases.";  
 RL J. Mol. Biol. 244:665-672 (1994).  
 RN [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=96178222; PubMed=8617202;  
 RA Colwill K., Pawson T., Andrews B., Prasad J., Manley J.L., Bell J.C.,  
 RT Duncan P.T.;  
 RL "The CLK/STY protein kinase phosphorylates SR splicing factors and  
 RT regulates their intranuclear distribution.";  
 RL EMBO J. 15:265-275 (1996).  
 CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS  
 CC OF THE SPLICING COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF  
 CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA  
 CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=P22518-1; Sequence=Displayed;

CC Name=Short;  
 CC IsoId=P22518-2; Sequence=VSP\_004854, VSP\_004855;  
 CC Note=Lacks the kinase domain.  
 CC -1- PFM: Autophosphorylates on all three types of residue.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X57186; CAA40473.1; -;  
 DR EMBL: M38381; AAA40151.1; -;  
 DR EMBL: U21209; AAC52257.1; -;  
 DR EMBL: L29221; AAA61485.1; -;  
 DR PIR: A39676; A39676.  
 DR PIR: I49275; I49275.  
 DR HSSP: Q00534; 1B18.  
 DR MGD: MGI:107403; CLK.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR ProDom: PD000001; Prot\_kinase; 1.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR ProSite: PS00107; PROTEIN KINASE\_ATP; 1.  
 DR ProSite: PS00108; PROTEIN KINASE\_ST; 1.  
 DR ProSite: PS50011; PROTEIN KINASE\_DOM; 1.  
 DR ProSite: PS50011; PROTEIN KINASE\_DOM; 1.  
 DR Tyrosinase; Serine/threonine-protein kinase; ATP-binding;  
 DR Tyrosine-protein kinase; Phosphorylation; Nuclear protein;  
 DR Alternative splicing.  
 DR DOMAIN 29 33 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 DR NP\_BIND 160 476 ATP (BY SIMILARITY).  
 DR BINDING 166 174 ATP (BY SIMILARITY).  
 DR ACT\_SITE 190 190 BY SIMILARITY.  
 DR VARSPLIC 130 287 KSHRRK -> MKLIL (in isoform Short).  
 DR FT FT Missing (in isoform Short).  
 DR VARSPLIC 136 483 /FTID=VSP\_004854.  
 DR FT FT /FTID=VSP\_004855.  
 DR MTTAGN 190 190 K-R: LOSS OF ACTIVITY.  
 DR CONFLICT 379 379 P -> S (IN REF. 2).  
 DR CONFLICT 448 448 F -> L (IN REF. 2).  
 DR CONFLICT 453 456 VKXI -> IGKM (IN REF. 2).  
 DR SQ SEQUENCE 483 AA; 57095 MW; 1D6BEC2AA2342619 CRC64;

Query Match 76.0%; Score 1832; DB 1; Length 483;  
 Best Local Similarity 79.7%; Pred. No. 1,2e-109; Indels 6; Gaps 4;  
 Matches 349; Conservative 32; Mismatches 51;

QY 9 HSEVEDTHPSHYEARSLENERDYDRRYVDEYNDYCEGVPHYRDIESGYRHCSS 68  
 DB 48 HSKTTD---SYLESSEINERKAVSRRYDEYNDYDM-GYEPHRYGEGSRYQWSSSS 103  
 QY 69 SVSRSSSPYRK-RNF-HCSSHQSRSKSHRRRSRSIEDDEGHILICSGDVIARYEIV 126  
 DB 104 SGRSGSSYSKSRSHRHSOHSRSHKSRHRRSRSEVEDDEGHILICSGDVIARYEIV 163  
 QY 127 DTLGEGAFGVVACIDHGDMGVAVKIKYKNGRYEARSSEIYVEHNSNDPNSVFC 186  
 DB 164 DTLGEGAFGVVACIDHGDMGVAVKIKYKNGRYEARSSEIYVEHNSNDPNSVFC 223  
 QY 187 VQMLEWFDRGHVACIVFELLGLSTYDFIKENSFLPFOIDHROMAYOICQINFLHNKL 246  
 DB 224 VQMLEWFDRGHVACIVFELLGLSTYDFIKENSFLPFOIDHROMAYOICQINFLHNKL 283  
 QY 247 THTDLKPENILFPKSDYVYVYKYSKMRDERTLKNDDIKYVDSGSAFYDDEHSTVSTR 306  
 DB 284 THTDLKPENILFPKSDYVYVYKYSKMRDERTLKNDDIKYVDSGSAFYDDEHSTVSTR 343



QY 307 YRAPEVILALGMSQPCDWSIGCIIIEYVIGFTVQTHDSKEHLAMERIIIGPIPOHMIQ 366  
 DB 344 YRAPEVILALGMSQPCDWSIGCIIIEYVIGFTVQTHDSKEHLAMERIIIGPIPOHMIQ 403  
 QY 367 KTRKRYEHNQDLMDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPVLVRMLEYDPTQ 426  
 DB 404 KTRKRYEHNQDLMDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPVLVRMLEYDPTQ 463  
 QY 427 RITLDEALQHPFDFLLKK 444  
 DB 464 RITLKEALKHPFFYPLKK 481

RESULT 4

CLK2\_MOUSE STANDARD; PRT; 499 AA.

ID CLK2\_MOUSE STANDARD; PRT; 499 AA.  
 AC 035491;  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protein kinase CLK2 (EC 2.7.1.-) (CDC-like kinase 2).  
 GN CLK2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=97439710; PubMed=9307018;  
 RA Nayler O., Stamm S., Ullrich A.;  
 RT "Characterization and comparison of four serine- and arginine-rich  
 RT (SR) protein kinases.";  
 RT Biochem. J. 326:693-700(1997).  
 RL Biochem. J. 326:693-700(1997).

CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS  
 CC OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF  
 CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA  
 CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).  
 CC -1- PTM: Autophosphorylates on all three types of residue.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

DR EMBL: A0033664; AAB87508.1; -.  
 DR MGI: 1098669; CLK2.  
 DR InterPro: IPR000719; Prot Kinase.  
 DR InterPro: IPR002290; Ser Thr kinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; Prot kinase; 1.  
 DR SMART: SM00220; S\_TKc\_1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 KW transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Tyrosine-protein kinase; Phosphorylation; Nuclear protein.  
 FT DOMAIN 163 479  
 FT NP\_BIND 169 177 ATP (BY SIMILARITY).  
 FT BINDING 192 192 ATP (BY SIMILARITY).  
 FT ACT\_SITE 289 289 BY SIMILARITY.  
 SQ SEQUENCE 499 AA; 59955 MW; 092946BA171C9ADD CRC64;

Query Match 59.2%; Score 1427.5; DB 1; Length 499;  
 Best Local Similarity 60.3%; Pred. No. 6.3e-84;  
 Matches 261; Conservative 69; Mismatches 96; Indels 7; Gaps 3;  
 18 SHYLBARS-----LNERDYRRRRYDEY-RNDYCEGYVRHRYHRIEISGYSRHCSSSVR 71

DB 50 SYHVSRSNYSYDHSDDLRYRRRCGSRRRNDYDRDGEAYVDDFROSLEYHNSYSR 109  
 QY 72 SRSS-PRKKNRHSQSSKSHRRKSSSIDDEBGLICQSGDYLRRAYEIVDTG 130  
 DB 110 SQRSSRRHRRRRRRSRFTFSRSSSHSSRRAKSVEDDAEGLIYHVGWLCRREYIVSLG 169  
 QY 131 EGAFKAYVECTDHQMDGHNVAVKIVKNVGRFREAARSIOYLEHANSIDPNSVFCVOML 190  
 DB 170 EGTSGRVVQCVDRHRRGGFRVALKTIKVEKYEKARLEINLEKNEQPNKMLCVOMF 229  
 QY 191 EWPDHGHVCIPELIGLSTYDFIKENSFLPQIDIRIOMAYQICQSNFLPHNKLTYTD 250  
 DB 230 DMFDYGHMCSFFELIGLSTDFPKDNNYLPYPIHQVRHAFQICQAYKFLHDKLTHD 289  
 QY 251 LKPENIIFVYSDYVYKYNKKKRDERTLKNIDIVVDFGSAATYDDEHHSITVSTRYAP 310  
 DB 290 LKPENIIFVYSDYVYKYNKKKRDERTLKNIDIVVDFGSAATYDDEHHSITVSTRYAP 349  
 QY 311 EVILALGMSQPCDWSIGCIIIEYVIGFTVQTHDSKEHLAMERIIIGPIPOHMIQ 370  
 DB 350 EVILALGMSQPCDWSIGCIIIEYVIGFTVQTHDSKEHLAMERIIIGPIPOHMIQ 409  
 QY 371 RKYFHNQDLMDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPVLVRMLEYDPTQ 430  
 DB 410 RKYFHNQDLMDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPVLVRMLEYDPTQ 469  
 QY 431 DEALQHPFDFLLKK 443  
 DB 470 DEALQHPFDFLLKK 482

RESULT 5

CLK2\_HUMAN STANDARD; PRT; 499 AA.

ID CLK2\_HUMAN STANDARD; PRT; 499 AA.  
 AC P49760; Q96CQ0;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protein kinase CLK2 (EC 2.7.1.-) (CDC-like kinase 2).  
 GN CLK2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=95082033; PubMed=7990150;  
 RA Hanes J.J., der Kammer H., Klaudiny J.J., Scheit K.H.;  
 RT "Characterization by cDNA cloning of two new human protein kinases.  
 RT Evidence by sequence comparison of a new family of mammalian protein  
 RT kinases.";  
 RT J. Mol. Biol. 244:665-672(1994).  
 RL J. Mol. Biol. 244:665-672(1994).

CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS  
 CC OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF  
 CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA  
 CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).  
 CC -1- PTM: Autophosphorylates on all three types of residue.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).

DR EMBL: A0033664; AAB87508.1; -.  
 DR MGI: 1098669; CLK2.  
 DR InterPro: IPR000719; Prot Kinase.  
 DR InterPro: IPR002290; Ser Thr kinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; Prot kinase; 1.  
 DR SMART: SM00220; S\_TKc\_1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 KW transferase; Serine/threonine-protein kinase; ATP-binding;  
 KW Tyrosine-protein kinase; Phosphorylation; Nuclear protein.  
 FT DOMAIN 163 479  
 FT NP\_BIND 169 177 ATP (BY SIMILARITY).  
 FT BINDING 192 192 ATP (BY SIMILARITY).  
 FT ACT\_SITE 289 289 BY SIMILARITY.  
 SQ SEQUENCE 499 AA; 59955 MW; 092946BA171C9ADD CRC64;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altshuler S.F., Zeeberg B., Buecaw K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong F.,  
 RA Brownstein M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Kana S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fandy J., Helton E., Kelleman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP FUNCTION.  
RX MEDLINE=97066903; PubMed=8910305;  
RT Lee K., Du C., Horn M., Rabinow L.,  
RT "Activity and autophosphorylation of LAMER protein kinases.";  
RT J. Biol. Chem. 271:27299-27303(1996).  
CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS  
CC OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF  
CC REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA  
CC SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES (BY  
CC SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long;  
CC IsoId=P49760-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=P49760-2; Sequence=VSP\_004856, VSP\_004857;  
CC -1- Note-lacks the kinase domain;  
CC -1- PTH: Autophosphorylates on all three types of residue.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC LAMER SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L29218; AAA61482.1; -;  
DR EMBL: L29216; AAA61481.1; -;  
DR EMBL: AF023268; AAC51817.1; -;  
DR EMBL: BC014067; AAL14067.1; -;  
DR PIR: S53637; S53637.  
DR PIR: S53638; S53638.  
DR Genew: HGNC:2069; CLK2.  
DR GK: P49760; -;  
DR MIM: 602989; -;  
DR GO: GO:0004674; P:protein serine/threonine kinase activity; TAS.  
DR InterPro: IPR000719; Prot. kinase.  
DR InterPro: IPR002290; Ser. Thr. kinase.  
DR Pfam: PF00069; Pkinase; 1.  
DR ProDom: PD000001; Prot. kinase; 1.  
DR SMART: SMO0220; S. TKC; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE: PS00114; PROTEIN\_KINASE\_DOM; 1.  
DR Translase: Serine/threonine-protein kinase; ATP-binding;  
RV Tyrosine-protein kinase; Phosphorylation; Nuclear protein;  
KV Alternative splicing.  
FT DOMAIN 163 479 PROTEIN KINASE.  
FT NP BIND 169 177 ATP (BY SIMILARITY).  
FT BINDING 193 193 ATP (BY SIMILARITY).  
FT ACT\_SITE 290 290 BY SIMILARITY.  
FT VASPLIC 134 139 QHSRR->VSKSLAP (in isoform short).  
FT VASPLIC 140 499 /Ftd-VSP\_004856.  
FT VASPLIC 140 499 Missing (in isoform short).  
FT VASPLIC 140 499 /Ftd-VSP\_004857.  
FT CONFLICT 134 134 MISSING (IN REF. 3).

SQ SEQUENCE 499 AA; 60090 MW; E43BFF3BAD6EE991 CRC64;  
Query Match 58.8%; Score 1417; DB 1; Length 499;  
Best Local Similarity 59.7%; Pred. No. 2.9e-83;  
Matches 259; Conservative 69; Mismatches 98; Indels 8; Gaps 3;  
QY 18 SHYLEARS-----LNEPRVDRRRYDEY-RNDYCEGVPRHVRHPIESGYRHCKSSVR 71  
DB 50 SYHVSRRSSYDRSSDRRYDRYRCGSTRANDYSRDGDYDTRYSRYEYGRSSSYR 109  
QY 72 SRRSPK-RKRNRRCSHQSRKSRKRRSRISDEDEGHLICSGVLRARVEYVDTL 129  
DB 110 SQRRSRKRRRRRRRRRRSRFRSSQHSRRRAKSVEDDAGHLIYHVGDLORRYEIVSTL 169  
QY 130 GEGARCKVYECIDHGMGMHVAVKYKVVGRREARAEIQVLEHNTDPSVRCVM 189  
DB 170 GGTTFRRVQCVDRHRCGARVVALKIKVKEKKEARLEINLEKINEKDPNNKVLCOVM 229  
QY 190 LEMFDHGHVCTVFELGLSTYDPTKENSFLPQIDHIROMAYQICQSIINFLHNKLTHT 249  
DB 230 FWFDFHGMCTSFELGLSTDFPKDNNYLPPIHQRRHMAFOLQAVKFLHDKLTHT 289  
QY 250 DKPENILFYKSDYVYKTRSKRKDEKTLKTDIVYDFGSGATYDDEHSTLVSRHRA 309  
DB 290 DKPENILFYNSDYELTYLLEKRDERSVKSVAHVDFGSGATYDDEHSTLVSRHRA 349  
QY 310 PVVIALGWSQPCDWSIGCILIEYLLGFTVFPQTHDSKEHLAMERILGPIQMHIOKTR 369  
DB 350 PEVILLEGMSQPCDWSIGCILIEYVGFILFQTHDNKEHLAMERILGPIPSRIIRTR 409  
QY 370 KKKYFHNDLWDDEHSSAGRYRRCKPLKEPMLCHDEHEKLPVLVRMLEYDPTQRT 429  
DB 410 KKKYFGRGLDWDENTSAGRYREKCKPLRLYLTSEAEHHQLFDLISMLEYPAKRLT 469  
QY 430 IDEALQHPFDLX 443  
DB 470 LGEALQHPFARLR 483  
RESULT 6  
CLNK HUMAN STANDARD; PRT: 490 AA.  
AC P49761; Q9BR53; Q9BU7;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Protein kinase CLNK (EC 2.7.1.-) (CDC-like kinase 3).  
GN CLNK3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=95082033; PubMed=7990150;  
RA Hanes J.J., der Kammer H., Klandou J.J., Scheit K.H.;  
RT "Characterization by cDNA cloning of two new human protein kinases:  
RT evidence by sequence comparison of a new family of mammalian protein  
RT kinases.";  
RT J. Mol. Biol. 244:665-672(1994).  
RN (2)  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).  
RC TISSUE=Cervix; and Placenta;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullani S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richard S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Witting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS  
 OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF  
 REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA  
 SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES (BY  
 SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1; Synonyms=Long;  
 CC Name=2; Synonyms=Short;  
 CC IsoId=P49761-2; Sequence=VSP\_004859, VSP\_004859;  
 CC Note=Lacks the kinase domain;  
 CC Name=3;  
 CC IsoId=P49761-3; Sequence=VSP\_004860;  
 CC -1- PTM: Autophosphorylates on all three types of residue (By  
 similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL: L29217; AAA61484.1; -;  
 DR EMBL: L29220; AAA61483.1; -;  
 DR EMBL: BC002555; AAA02555.1; -;  
 DR EMBL: BC006103; AAH06103.1; -;  
 DR EMBL: BC019881; AAH19881.1; -;  
 DR PIR: S53639; S53639.  
 DR PIR: S53640; S53640.  
 DR HSSP: Q00534; 1B18.  
 DR Genew: HGNC:2071; CLK3.  
 DR GK: P49761; -;  
 DR MTM: 602990; -;  
 DR GO: GO:000474; F1protein serine/threonine kinase activity; TAS.  
 DR InterPro: IPR000719; Prot\_Kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; Prot\_kinase; 1.  
 DR SMART: SM00220; S\_TKc.1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
 DR PROSITE: PS00111; PROTEIN KINASE\_D0M; 1.  
 DR Transfease: Serine/threonine-protein kinase; ATP-binding;  
 KW Tyrosine-protein kinase; Phosphorylation; Nuclear protein;  
 KW alternative splicing.  
 FT DOMAIN 156 472 PROTEIN KINASE.  
 FT NP\_BIND 162 170 ATP (BY SIMILARITY).  
 FT BINDING 186 186 ATP (BY SIMILARITY).  
 FT ACT\_SITE 283 283 BY SIMILARITY.  
 FT VARSPIC 124 152 RSQSSKRSRSVEDDEKGLHLCVIGDML -> MELMGWTW  
 FT VARSPIC 153 430 /FTId=VSP\_004858.  
 FT VARSPIC 158 180 /FTId=VSP\_004859.  
 FT VARSPIC 158 180 /FTId=VSP\_004860.  
 FT VARSPIC 158 180 /FTId=VSP\_004860.

FT CONFLICT 132 133 SS -> TG (IN REF. 1).  
 SQ SEQUENCE 490 AA; 56588 MW; 86A342AB8AB24CA CRC64;  
 Query Match 54.8%; Score 1321; DB 1; Length 490;  
 Best Local Similarity 55.3%; Pred. No. 3, 6e-77;  
 Matches 251; Conservative 60; Mismatches 101; Indels 42; Gaps 6;  
 -----  
 QY 4 PLEASHSVEEDTHPSHYLEARSINERDYDRDYDEYNDYCGYVP-----R 51  
 DB 43 PPRRSRSRSDRLP-----YORRYRRRSDTYR---CERSPSPGDEYDGPSSR 90  
 QY 52 HYHDDISG-YRI-----HCKRSVSRRESRPRKKNHSHQSSKSHRRRSRIED 105  
 DB 91 HRRRSRERGYYRKRKAHCHK-----RTRSCSSASSSQSSKSSSVED 138  
 QY 106 DEEGHLICQSGDYLRAEYELVDTLGEGAFGKVEICIDHGMDGMVAVKYKNGRYREA 165  
 DB 139 DKEGHLVCRIGDMLQERIEYIVGNLGGTFFGKVECDLHARGSQVALKIRANGKYREA 198  
 QY 166 RSEIQLVHLNSTDNPSVFCVQMLMEFHDHGHVCIYFELIGLSTYDFIKENSFLPQID 225  
 DB 199 RLEINVLAKKIKKDKENKFLCVLMSDFNFGHMCIAPELLGKNTPEFLKNNFQYPLP 258  
 QY 226 HIROMAYQICQSGINFLHNKLTHTDLPENILFVKSDYVVKYKSKKRDERTLKNTDIKY 285  
 DB 259 HVRMAYQICHALRFLHENGULTHTDLPENILFVNSEPEFLYNHKSCEKSVKNTSIRY 318  
 QY 286 VDRGSATYDDEHNSLTVSTHYPAPVYLALGMSQCDVWSICGILIEYLGTVQTHD 345  
 DB 319 ADFGSATFDEHNTHTVATRHYPPEVILELGAQPCDVWSICGILIEYRGTLTQTHE 378  
 QY 346 SKEHLAMERILIPIPOMHIQKRRKRYFHNLDMDEHSSAGRYRRCKPLKEMLCH 405  
 DB 379 NREHLVMEKILIPISPHMIRKOKYFKGGLVNDENSDRKYKENEKPLKSYTLQD 438  
 QY 406 DEEHKLPDLVRMLBYDPTQRTILDALQHPFF 439  
 DB 439 SLEHVQLFLMRRLMEFDPQRITTLBALHPFF 472  
 -----  
 RESULT 7  
 CLK3 RAT STANDARD; PRT; 490 AA.  
 ID CLK3 RAT  
 AC Q63117;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protein kinase CLK3 (EC 2.7.1.-) (CDC-like kinase 3).  
 GN CLK3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain.  
 EX MEDLINE=96271481; PubMed=8679717;  
 RA Becker W., Kentrup H., Henkelbach U., Joost H.G.;  
 RT "cDNA cloning and characterization of rat CLK3, a LAMMER kinase  
 predominantly expressed in testis.";  
 RL Biochim. Biophys. Acta 1312:63-67(1996).  
 CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGININE-RICH (SR) PROTEINS  
 OF THE SPliceosomal COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF  
 REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA  
 SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES (BY  
 SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- PTM: Autophosphorylates on all three types of residue.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X8351; CAA64076.1; --

DR PIR; S70352; S70352.

DR HSSP; Q00534; 1B18.

DR InterPro; IPR000719; Prot. kinase.

DR InterPro; IPR002290; Ser. thr. kinase.

DR Pfam; PF00069; pkinase; 1.

DR ProDom; PD000001; Prot. kinase; 1.

DR SMART; SM00220; S TKC; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS00108; PROTEIN KINASE ST; 1.

DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.

DR Transferrase; Serine/threonine-protein kinase; ATP-binding; tyrosine-protein kinase; Phosphorylation; Nuclear protein.

FT DOMAIN 156 472

FT NP BIND 162 170

FT BINDING 186 186

FT ACT SITE 283 283

FT SEQUENCE 490 AA; 58501 MW; 72F754F6A64B19A CRC64;

Query Match 54.6%; Score 1317; DB 1; Length 490;

Best Local Similarity 55.1%; Pred. No. 6.6e-77;

Matches 250; Conservative 61; Mismatches 101; Indels 42; Gaps 6;

4 PLEASHSVEDTHPHYLKASLSNEDRYDEYNDCEGVY-----R 51

43 PRRSRSSSHDRIP-----YORRYRHRDSDTYR--CEERSSPFGEDCYGSSRSR 90

52 HYHRDIEEG-YRI-----HCSKSSVSRSSSPKRRNHCSSHQSRKSHRRKRSIED 105

91 HRRRSRERGYPYTRKHAHCHK-----RRTSSGSSASSRSGQSSKRSRVED 138

106 DEEGHLCQSGDVLPAKYEIYDTLGEAGFGKVEYEDIDGMGMAVAKYKNGRYEAA 165

139 DEEGHLCQSGDVLPAKYEIYDTLGEAGFGKVEYEDIDGMGMAVAKYKNGRYEAA 198

166 RSEIQLVLEHNTDENSVCQMLEWFDHGHVCIYFELLGLSTYDFIKENSPFPQID 225

199 REIIVLVKIKKIKKDKENKFLCVLMDWENPHGHMCIAPELLKNTPELKENNQPPPLP 258

226 HIRQMAVQICQSLNHLNKLTHDLPKENVLFVKSADYVYKNSKMRDEKTLKNDIKY 285

259 HVRHMAVQICQSLNHLNKLTHDLPKENVLFVKSADYVYKNSKMRDEKTLKNDIKY 318

286 VDFGSATYDDEHSLVSTRYARAEVILALGMSQPCVWSIGCILAEYLGFTVPTD 345

319 ADFGSATYDDEHSLVSTRYARAEVILALGMSQPCVWSIGCILAEYLGFTVPTD 378

346 STEHLAMMERILGIPQHYIOTKRRKFFHNQDMDHSSAGRYVRCKPLKFMELCH 405

379 NNEHVMMEKILGIPSHIHTRKOKYFVKGGLVWENSSDGRYVKNCKPLKSYMLQD 438

406 DEHEKLFPLVRYRMLEYDPTGRTTDLDELQHPFF 439

439 STEHQLFDLMKRMLEFPDSQRTITAEALHPPF 472

RESULT 8

CLK3 MOUSE STANDARD; PRT; 490 AA.

AC 035432;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Protein kinase CLK3 (EC 2.7.1.1.-) (CDC-like kinase 3).

OS CLK3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Mus.

NCBI\_TaxID=10090;

11

SEQUENCE FROM N.A.

Medline=97439710; PubMed=9307018;

RA Naylor O., Stamm S., Ullrich A.;

RT "Characterization and comparison of four serine- and arginine-rich (SR) protein kinases."

RL Biochem. J. 326:693-700(1997).

CC -1- FUNCTION: PHOSPHORYLATES SERINE- AND ARGinine-RICH (SR) PROTEINS OF THE SPICEOSOMAL COMPLEX MAY BE A CONSTITUENT OF A NETWORK OF REGULATORY MECHANISMS THAT ENABLE SR PROTEINS TO CONTROL RNA SPLICING. PHOSPHORYLATES SERINES, THREONINES AND TYROSINES.

CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).

CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN TESTIS.

CC -1- PTM: Autophosphorylates on all three types of residue.

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. LAMMER SUBFAMILY.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF033565; AAB97509.1; --

DR HSSP; Q00534; 1B18.

DR MD; MG1:1098670; CLK3.

DR InterPro; IPR000719; Prot. kinase.

DR InterPro; IPR002290; Ser. thr. kinase.

DR Pfam; PF00069; pkinase; 1.

DR ProDom; PD000001; Prot. kinase; 1.

DR SMART; SM00220; S TKC; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE; PS00108; PROTEIN KINASE ST; 1.

DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.

DR Transferrase; Serine/threonine-protein kinase; ATP-binding; tyrosine-protein kinase; Phosphorylation; Nuclear protein.

FT DOMAIN 156 472

FT NP BIND 162 170

FT BINDING 186 186

FT ACT SITE 283 283

FT SEQUENCE 490 AA; 58513 MW; A7C0311C530AC41 CRC64;

Query Match 54.4%; Score 1310; DB 1; Length 490;

Best Local Similarity 55.6%; Pred. No. 1.8e-76;

Matches 248; Conservative 62; Mismatches 110; Indels 26; Gaps 4;

4 PLEASHSVEDTHPHYLKASLSNEDRYDEYNDCEGVYPRHHRDIEGRI 63

43 PRRSRSSSHDRIP-----YORRYRHRDSDTYR--CEERSSPFG---EDCYGS 86

64 HCSKSSVSRSSSPKRRNHCSSHQSRKSHRRKRSIEDDEGHLC 113

87 SRSRRHRRSRERAPYTRKHAHCHKRTRSCSSASSRSGQSSKRSRVEDDEGHLYC 146

114 QSGDVLPAKYEIYDTLGEAGFGKVEYEDIDGMGMAVAKYKNGRYEAAEIQYLE 173

147 RIGDWLQERYEIVGLGEGTGFQVCELDHARGSOVALKIRNKGKREARLEIVLK 206

174 HLNSTDPSNVSFCVQMLEWFDHGHVCIYFELLGLSTYDFIKENSPFPQIDHROMAYQ 233

207 KIKEDKDKENKFLCVLMDWENPHGHMCIAPELLKNTPELKENNQPPPLPVRHMAVQ 266

234 IQSLNHLNKLTHDLPKENVLFVKSADYVYKNSKMRDEKTLKNDIKYVDFGATY 293

267 LCHALRFHENDLTHDLPKENVLFVNSEFTLVNHCSEKSKVNTSIRVADFGATF 326

294 DDEHSLVSTRYARAEVILALGMSQPCVWSIGCILAEYLGFTVPTDSEKHLAM 353

327 DHEHTTIVATRHRRPREVILDELMAQPCVWSIGCILAEYLRGFTLQTHENREHLVM 386

QY 354 ERICGPIPOHMTQTRKRTKPFHNDLMDHSSAGRYVRRCPLKPMCHDEHEKLP 413  
 DB 387 EKILGPIPSHMTHTKRTKPFYKGLVMDNSSDGRVYKENCPLKMSYMLQSLHEVOLF 446  
 QY 414 DLVRMLEYDPTORTLDEALQHPPE 439  
 DB 447 DLVRMLEYDPTORTLDEALQHPPE 472

RESULT 9  
 ID DOA DROME STANDARD; PRT; 517 AA.

AC P49762: O9VAR8;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protein kinase Doa (EC 2.7.1.-) (Protein darkener of apricot).  
 GN DOA OR CG1658.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CK NCBI\_TaxID=7227;

RP SEQUENCE FROM N.A.  
 RX MEDLINE=95011531; PubMed=7926721;  
 RA Yun B., Parkas R., Lee K., Rabinow L.;  
 RT "The Doa locus encodes a member of a new protein kinase family and is  
 RT essential for eye and embryonic development in Drosophila  
 RT melanogaster.";  
 RL Genes Dev. 8:1160-1173(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Rabinow L.;  
 RN Submitted (JUN-1994) to the EMBL/GenBank/DBD databases.

RP [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA Sutton G.G., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutcliffe G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champ M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Berens P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
 RA Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,  
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,  
 RA Jaitai M., Kalish F., Kapran G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon S., Nusskern D.R., Paclet J.M.,  
 RA Palazolo M., Pittman G.S., Fan S., Pollard V., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).

RP [4]  
 RP FUNCTION, AND MUTAGENESIS OF LYS-199.  
 RX MEDLINE=97066903; PubMed=8910305;  
 RA Lee K., Du C., Horn M., Rabinow L.;  
 RT "Activity and autophosphorylation of LAMMER protein kinases.";  
 RL J. Biol. Chem. 271:27299-27303(1996).

CC - FUNCTION: Negative regulator of the copia retrotransposon element  
 CC of the white (w) gene. In the eye, it is required for normal  
 CC pigmentation, photoreceptor cell development and for organization  
 CC of interommatidial bristles. Also essential for embryonic  
 CC segmentation and differentiation of the nervous system.

CC - DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT DEVELOPMENT.  
 CC - SIMILARITY: Autophosphorylates on all three types of residue.  
 CC - SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.

CC - CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-7 IS THE INITIATOR.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: X78715; CAA5367.1; -  
 CC EMBL: AE003767; AAF6832.2; ALT\_INIT.  
 CC PIR: A54099; A54099.  
 CC FlyBase: FBgn0000480; Doa.  
 CC GO: GO:0005737; C:cytoplasm; IDA.  
 CC GO: GO:0005634; C:nucleus; IDA.  
 CC GO: GO:0004674; F:protein serine/threonine kinase activity; IDA.  
 CC GO: GO:0007350; P:biostoderm segmentation; IMP.  
 CC GO: GO:0042051; P:eye photoreceptor development (sensu Drosop. . .); IEP.  
 CC GO: GO:0007399; P:neurogenesis; IMP.  
 CC GO: GO:0007548; P:sex differentiation; IGI.  
 CC InterPro: IPR000719; Prot\_kinase.  
 CC InterPro: IPR002290; Ser\_thr\_kinase.  
 CC Pfam: PF00069; Pkinase; 1.  
 CC ProDom: PD000001; Prot\_kinase; 1.  
 CC SMART: SM00220; S\_TKC; 1.

CC PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 CC PROSITE: PS00109; PROTEIN KINASE ST; 1.  
 CC PROSITE: PS00111; PROTEIN KINASE DOM; 1.  
 CC Transferrase, Serine/threonine-protein kinase, ATP-binding;  
 CC Tyrosine-protein kinase; Phosphorylation.  
 CC DOMAIN 170 484  
 CC NP\_BIND 176 184  
 CC BINDING 199 199  
 CC ACT\_SITE 296 296  
 CC MUTAGEN 199 199  
 CC SEQUENCE 517 AA; 60975 MW; 830EB8A950CEDA1 CRC64;

Query Match 50.7%; Score 1222; DB 1; Length 517;  
 Best Local Similarity 57.8%; Pred. No. 7.6e-71;  
 Matches 230; Conservative 58; Mismatches 84; Indels 26; Gaps 4;

QY 69 SVRRSSRRKRRKRRHSSHO---SRKSRRRKRSR-----IE 104  
 DB 92 SKRSRRSSPAASRRKRYRDETSSSSRRRDRKDRSDGRNRQSOAKTAPVQ 151  
 QY 105 DDEGHGICGSDVLRARAYELVDTLGGAGCKVCEIDHGDMGVAVKTVKNGRYREA 164  
 DB 152 DDADGHLIYHGDIILHRYKIMATLGEGTGRVVKVDMERD-YCALKILIIKVEKXREA 210  
 QY 165 ARSEQLVLEHNSDPSVRCQMLEPFGHGVICVFELLGLSTYDFIKENSFLPQI 224  
 DB 211 AKELINALEKIAQDPRCDHLCYAMIDMPEYHGMCIVFEMLGISVDFDRENNYEPYL 270



QY 225 DHIRQMAVYOCOSINFLHNKLTHTDLKPEENLFYKSYDVVKYNSMKMKDEKTLKNTDITK 284  
 DB 271 DQVFNHAYOUCYSKFLADNRLLHTDLKPEENLIFVSDTSYSHNKHINEVVRKNTDVR 330  
 QY 285 VVDEGSATYDEHNSHTVSTRHRAPEVILALGMSOPCVMSIGCTILIEYVGFVFPOT 344  
 DB 331 LIDGSAFTPEHNSHTVSTRHRAPEVILALGMSOPCVMSIGCTILIEYVGFVFPOT 390  
 QY 345 DSKENHLMERIIIGPDPQHNITQTRKRRKFFHNOLDWDEHSSAGRYVRRCKPLKFMIC 404  
 DB 391 DNEHLMAMERIIIGPDPQHNITQTRKRRKFFHNOLDWDEHSSAGRYVRRCKPLKFMIC 449  
 QY 405 HDEHEKLPDLYVRMLEYDPTORTITLDEALQHPFDLT 442  
 DB 450 DSEHCEHFLSKMLKMLPEPSRITLGEALHMFDFDL 487  
 RESULT 10  
 AF02\_ARATH STANDARD; PRT; 427 AA.  
 ID AF02\_ARATH  
 AC P51567  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein kinase AF02 (BC 2.7.1.-).  
 GN AF02 OR AT4G24740 OR F22K18.60.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eusteroideae; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Landsberg erecta;  
 RX MEDLINE=95083650; PubMed=7991592;  
 RA Bender J., Fink G.R.;  
 RT "ARCT1, a LAMMER kinase from Arabidopsis thaliana, activates STE12-  
 dependent processes in yeast.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:12105-12109(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Kuromori T., Yamamoto M.;  
 RT "A thaliana gene encoding protein kinases of a new family.";  
 RL Submitted (FE8-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Schaeffler A., Stiekema W., Entian K.-D., Terry N.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,  
 RA Harris B., Ansoorge W., Brandt P., Grievell L.A., Rieger M., Mueller M.,  
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,  
 RA Kreis W., Delseny M., Puigdomenech P., Watson M., Schmidtnehl T.,  
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hohselt U., Zimmermann W., Wedler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA Van der Schueren T., Gymnoperez B., Chuang Y.-J., Vandenbussche F.,  
 RA Braeken M., Welljens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weitzsaecker T., Bothe G., Rampsperger U., Hilbert H., Braun M.,  
 RA Holzner E., Brandt A., Peters S., Van Staaveren M., Dirke W.,  
 RA Woolfjan P., Klein Lankehorst R., Rose M., Hauf J., Koester P.,  
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 RA De Keyser A., Buyschaert C., Giesen J., Villarroel R., De Clercq R.,  
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K., Mayes R.,  
 RA Petter A., Rajandream M.A., Lyne M., Benes V., Rehmann S.,  
 RA Borkova D., Blocker H., Scharfe M., Grimm M., Lohmeyer T.-H.,  
 RA Dose S., de Haan W., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzi A.,  
 RA Messner O., Aigley A., Vitale D., Lignocri R., Piravandi E.,  
 RA Messner O., Aigley A., Vitale D., Lignocri R., Piravandi E.,  
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,

RA Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandembol M., Barquas M., Terol J., Torres A.,  
 RA Perez-Perez A., Punelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Hejlskov L., Schwarz S., Scholler P., Heber S., Francs P., Biele C.,  
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Beyan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Pannell L., Dedbia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh U.,  
 RA Stoneking T., Kallik J., Graves T., Harmon G., Edwards J.,  
 RA Latreille P., Courtney L., Clond J., Abbott A., Scott K., Johnson D.,  
 RA Mink P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kremer J., Fulton B., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Spieghel J., Ryan E., Andrews S., Giesel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones K., Dyrce K., Cotton M., Joshi C.,  
 RA Antoniou B., Zidan M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Wa P., Zhong J., Preston R., Vill D., Shekhar M., Matro A., Shah R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,  
 RA Granat S., Shohdy N., Hasegawa A., Rodriguez M., Hoffman J., Tili S.,  
 RA Chen B., Marra M., Martienssen R., McCombie W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 402:769-777(1999).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U16177; AAA57118.1; -;  
 DR EMBL; D45353; BAA08214.1; -;  
 DR EMBL; AL035356; CA822989.1; -;  
 DR EMBL; AL161562; CA879384.1; -;  
 DR PIR; T05560; T05560.  
 DR HSSP; P24941; HCL.  
 DR InterPro; IPR000719; Prot kinase.  
 DR InterPro; IPR002290; Ser Thr kinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR ProDom; PD000001; Prot kinase; 1.  
 DR SMART; SMO0220; 5\_TKC1; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP, FALSE\_NEG.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR TRANSFERASE; Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 98 423  
 FT NP BIND 104 112 ATP (BY SIMILARITY).  
 FT BINDING 127 127 ATP (BY SIMILARITY).  
 FT ACT\_SITE 223 223 BY SIMILARITY.  
 SQ SEQUENCE 427 AA; 49566 MW; 33A1948B8372B374 CRC64;  
 Query Match 35.2%; Score 848.5; DB 1; Length 427;  
 Best Local Similarity 48.1%; Pred. No. 3.5e-47;  
 Matches 167; Conservative 70; Mismatches 97; Indels 13; Gaps 8;  
 QY 104 EDEEGLHLCQSGDVAFARIEYDITGEGAFGRVYECIDHMDGHAAYAKYKNGRYE 163  
 DB 79 EDDDGHYIELGDDTLPRKYIKSKGEGFGVGLCWDERKEMVAAYIVRGVKKYE 137  
 QY 164 AARSEIOLVHLNSTDPNSVFCVOMLEMDHGHGVCIVELGLSTYFIKENSFLPFO 223  
 DB 138 AAMEIEMLDQGLGHDGCG-NRCVQIRNMDYNNHICIVFEKGSLLYFLRKNVRSFP 196  
 QY 224 IDHRCQAVYOCOSINFLHNKLTHTDLKPEENLFYKSYDV-VK-NSMKMD--ERT 277  
 DB 197 IDLVRBEGWDLICVAFMDLHMTDLKPEENLIVSSDYVKIPEYKGSRLQEDVCYKEV 256  
 QY 278 LKNDIVVFGSATYDEHNSHTVSTRHRAPEVILALGMSOPCVMSIGCTILIEYVGF 337  
 DB 257 PKSAIVIVDFSGTTRRQDQYIVSTRHRAPEVILALGMSOPCVMSIGCTILIEYVGF 316

QY 338 FVFPQTHSKETLAMEERLGPIDPMIOKT--RKKYFHNOLDWDEHSSAGRYRRRC 395  
 Db 317 EALFQTHNLEHLMAMERVLGPPQOMKKVRHSEKTVRRRCRLDMPGARSRLKXVL 376  
 QY 396 K--PLKEFPLCH-DEEHEKLPDVRMLLEYDPTQITLDEALQHPFF 439  
 Db 377 KLPRLQNLIMQVHDSAGELINNOGLRFPDSEIRTAERLHPFF 423

RESULT 11  
 AFCL ARATH STANDARD; PRT; 467 AA.  
 AC P51568; Q39185; (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protein kinase AFCL (EC 2.7.1.1).  
 GN AFCL OR AME2 OR ATG53570 OR FAP12.270.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
 OK NCBI\_TaxId=3702;  
 RN 1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Landsberg erecta;  
 RX MEDLINE=95083650; PubMed=7991592;  
 RA Bender J., Fink G.R.;  
 RT "AFCL, a LAMMER kinase from Arabidopsis thaliana, activates STE12-  
 dependent processes in yeast."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:12105-12109(1994).  
 RN 2)  
 RP SEQUENCE FROM N.A.  
 RA Kurumori T., Yamamoto M.;  
 RT "A thaliana genes encoding protein kinases of a new family."  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 RN 3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016720; PubMed=1130713;  
 RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unsel M.,  
 RA Farmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 RA Deleney M., Boutry M., Artveit L.A., Mache R., Puigdomenech P.,  
 RA De Simone V., Choisme N., Gillemeau F., Robert C., Broclet P.,  
 RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quelier F.,  
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 RA Wurbach E., Drzonek H., Erle H., Jordan N., Bangert S.,  
 RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,  
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppe S., Simonietti B.,  
 RA Conrad A., Hornischer K., Kauer G., Boehmer T.-H., Nordiek G.,  
 RA Reichelt U., Schalte M., Schen O., Barques M., Terol J., Clement J.,  
 RA Navarro P., Collado C., Perez-Perez A., Oltgenwelder B., Duchemin D.,  
 RA Coore R., Lande M., Berger-Liauro C., Purnelle B., Masny D.,  
 RA De Haan M., Maere A.C., Alcaraz J.-P., Cortet A., Casacuberta E.,  
 RA Montfort A., Argitrou A., Flores M., Liguori R., Vitale D.,  
 RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mees H.-W.,  
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 RA Rooney T., Rizzo M., Walts A., Uteirack T., Fujii C.Y., Shea T.P.,  
 RA Cressy T.H., Haas B., Malet R., Wu D., Peterson J., Van Aken S.,  
 RA Pai G., Miltischer J., Sellers P., Gill J.E., Feldblum T.V.,  
 RA Preuss D., Lin X., Niernan M.C., Salzberg S.L., White O., Venter J.C.,  
 RA Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,  
 RA Matsunabe A., Yamada M., Yasuda M., Tabata S.;  
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 thaliana."  
 RL Nature 408:820-822(2000).  
 CC -1- FUNCTION: ACTIVATOR OF YEAST TRANSCRIPTION FACTOR, STE12.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U16176; AAA57117.1; -  
 DR EMBL; D45354; BAA08215.1; -  
 DR EMBL; AL132966; CAB67664.1; -  
 DR PIR; S71169; S71169.  
 DR HSBP; P24941; 1A01.  
 DR InterPro; IPR000719; Prot kinase.  
 DR InterPro; IPR002290; Ser thr kinase.  
 DR InterPro; IPR001245; Tyr kinase.  
 DR Pfam; PF00069; kinase; 1.  
 DR Prodom; PD000001; Prot kinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP, FALSE\_NEG.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transferrase; Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 115 443 PROTEIN\_KINASE.  
 FT NP\_BIND 121 129 ATP (BY SIMILARITY).  
 FT BINDING 144 144 ATP (BY SIMILARITY).  
 FT ACT\_SITE 240 240 BY SIMILARITY.  
 FT CONFLICT 117 117 I -> M (IN REF. 1).  
 SQ SEQUENCE 467 AA; 54198 MW; A885FD32CE11B181 CRC64;

Query Match 34.4%; Score 830; DB 1; Length 467;  
 Best Local Similarity 45.8%; Pred. No. 5.7e-46;

Matches 160; Conservative 71; Mismatches 102; Indels 16; Gaps 8;

QY 105 DDEEGHLCQSGGVLPARYEIVDTLGEAGFGKVECTIDHGMQMVAVKIVNNGRYREA 164  
 Db 97 DDKDGHVFEVGGDTLPRYQILSKMGEGFGVLECFDMKNREV-VAIKVIRINKYRBA 155  
 QY 165 ARSEIQLVLEHNTSDNSVFCVOMLEMDHGHVIVELGLSTYDITKENSFLPQI 224  
 Db 156 AMEIVDLQRLTRHDVG-SRCVQIKRMFPYRNHITVEKLGPSLYDLRKNYSRSP 214  
 QY 225 DTRQMAVOICQISINFLHNKLTHTDLPKENTLFEVSDV---VKYNSKMRDEKTLN 280  
 Db 215 DLVRELGRQLSEVAVMHDLRIHDLKPEMLIVSSEYIKIPDYKFLSRPTKDSYFNK 274  
 QY 281 ----TQIKVDFRSATYDDEHSTLVSTRHYAPVYILALGMSQPCDWSIGCLILEYL 336  
 Db 275 LPRSSAIKLIDFSTTFEHDHNYIVSTRHYAPVYILGVMVPCDLSIGCLILELCS 334  
 QY 337 GFVFPQTHSKETLAMEERLGPIDPMIOKT--RKKYFHNH-OLDWDEHSSAGRYRR 393  
 Db 335 GEAFLPQTHNLEHLMAMERVLGPPQOMKKVRHSEKTVRRRCRLDMPGARSRLKXVL 394  
 QY 394 --RCKPLKEFPLCH-DEEHEKLPDVRMLLEYDPTQITLDEALQHPFF 439  
 Db 395 VWKLPLRLQNLIMQVHDSAGELIDLOGLRLRYDPTERFARALNHPFF 443

RESULT 12  
 AFCL ARATH STANDARD; PRT; 400 AA.  
 AC P51568; Q39185; (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 40, Last annotation update)  
 DE Protein kinase AFCL (EC 2.7.1.1).  
 GN AFCL OR AME2 OR ATG53266 OR FAP11.140.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;



CC eucoside II; Brassicales; Brassicaceae; Arabidopsis.  
 CC NCBI\_TaxID=3702;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC Kurumori T., Yamamoto M.;  
 CC "A. thaliana genes encoding protein kinases of a new family.";  
 CC Submitted (Feb-1995) to the EMBL/GenBank/DBJ databases.  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=cv. Columbia;  
 CC MEDLINE=20083488; PubMed=10617198;  
 CC Mayer K.F.X., Schueller C., Wandt R., Murphy G., Volckaert G.,  
 CC Pohl T., Duesterhoef A., Striekema W., Entian K.-D., Terry N.,  
 CC Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,  
 CC Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,  
 CC Kreis M., Delseny M., Flugdonenach P., Watson M., Schmidtmann T.,  
 CC Reichert B., Portelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 CC Vos P., Hobeisel U., Zimmermann W., Medler H., Ridley P.,  
 CC Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 CC Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandebussche F.,  
 CC Breken M., Weltens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 CC Weitzneger T., Bothe G., Rampsberger U., Hilbert H., Braun M.,  
 CC Holzner E., Brandt A., Peters S., van Staveren M., Dirke W.,  
 CC Woolman P., Klein Lanhorst R., Rose M., Hauf U., Koetter P.,  
 CC Berneser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
 CC De Keyser A., Buyschaert C., Giesen J., Villarroel R., De Clercq R.,  
 CC Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,  
 CC Clark L., Doggett J., Hall S., Kay M., Leonard N., McIay K., Mayes R.,  
 CC Petter A., Rajendram M.A., Lyne M., Benes V., Reckmann T.-H.,  
 CC Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.,  
 CC Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 CC Gabel C., Fuchs W., Fartmann B., Grandtath K., Dauner D., Herzl A.,  
 CC Neumann S., Argitlon A., Vitale D., Liguori R., Pitarova E.,  
 CC Massenot O., Quigley F., Cladaud G., Muendlein A., Felber R.,  
 CC Schnal S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
 CC Cheffor F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
 CC Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 CC Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 CC Hejlsman D., Schwarz S., Scholler P., Heber S., Francis P., Bialke C.,  
 CC Fjstman L., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 CC Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 CC Pannell L., Dehila N., Gnoj L., Schutz K., Huang E., Siegel L.,  
 CC Sehon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 CC Steneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 CC Lattelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 CC Mix P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 CC Kramer J., Fulton L., Matis B., Dante W., Pepin K., Hillier L.,  
 CC Nelson J., Splich U., Ryan E., Andrews S., Geisel C., Layman D.,  
 CC Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
 CC Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,  
 CC Ma P., Zhong J., Preston R., Vil D., Sheker M., Matero A., Shah R.,  
 CC Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 CC Granal S., Shohay N., Hasegawa A., Hamed A., Lodi M., Johnson A.,  
 CC Chen E., Marra M., Martensen R., McCormie W.R.;  
 CC "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 CC thaliana.";  
 CC Nature 402:769-777(1999).  
 CC [3]  
 CC SEQUENCE OF 6-400 FROM N.A.  
 CC STRAIN=cv. Landsberg erecta;  
 CC MEDLINE=95083650; PubMed=7991592;  
 CC Bender J., Fink G.R.;  
 CC "AFCL, a LAMMER kinase from Arabidopsis thaliana, activates STE12-  
 CC dependent processes in yeast.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 91:12105-12109(1994).  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC LAMMER SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; D45355; BA08216.1; -  
 CC DR EMBL; AL022537; CAA18595.1; -  
 CC DR EMBL; AL161581; CAB79983.1; -  
 CC DR EMBL; U16178; AA57119.1; -  
 CC DR PIR; T04460; T04460.  
 CC DR HSSP; P24941; IAO1.  
 CC DR InterPro; IPR000719; Prot\_kinase.  
 CC DR InterPro; IPR002290; Ser\_thr\_kinase.  
 CC DR Pfam; PF00069; Pkinase; 1.  
 CC DR Prodom; PD000001; Prot\_kinase; 1.  
 CC DR SMART; SM00220; S\_TKC\_1.  
 CC DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 CC DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 CC DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC DR Transferrase; Serine/threonine-protein kinase; ATP-binding.  
 CC FT DOMAIN 71 396  
 CC NP\_BIND 77 85  
 CC BINDING 100 100  
 CC ACT\_SITE 196 196  
 CC FT COMPACT 8 8  
 CC FT SEQUENCE 400 AA; 46224 MW; 8F707ABDCFB1352 CRC64;  
 CC Query Match 32.8%; Score 790; DB 1; Length 400;  
 CC Best Local Similarity 41.7%; Pred. No. 1.7e-43;  
 CC Matches 169; Conservative 70; Mismatches 134; Indels 32; Gaps 11;  
 CC  
 CC 57 IESGRIHCKSSVRS-----RSPPKR--NRHGSQSSKSHRRKRSSTED 105  
 CC 2 IANGFE-SMDKRRKPRMTWDEAPPEAKRAVKGSGDRIISPPLR-----DD 53  
 CC 106 DEEGHLLCGSGDGVRAAREYIVDTLGEAFKVEGCEIDHGMDGHWAVKIYKNGRYREA 165  
 CC 54 DRDHVYFSKRDUTPRYKILSKMGSGTFRVLECDMDPTK-EYVAIKIIRSKIKRDA 112  
 CC 166 RSEIVLEHLNSTDPNSVFCVOMLEWPDHGHGVCIVFELLGSTYDFIKENSFLPPQID 225  
 CC 113 MIEDIVLOKLVKSKDKRT-RCVOMKWFDRNHICIVFEKLGSLDFELKKNKSAFPLA 171  
 CC 226 HIRGMAVOIGOSINFLHNHNLTPDILKPEVILFVKSQDVYVKNKSKKIDERT-----LKN 280  
 CC 172 LVRFPGGLLESAVYMELOLVHTDUKEPILVSEENVKLPNKKSSANETFRCLPKS 231  
 CC 281 TDIVVDFSGATYDDHSHSTLVSTRYAPDEVITALGWSOPCDVWSIGCILIEYVIGFTV 340  
 CC 232 SAKILDFGSGTVCNRIHSHIVOTRHYSPEVILGWSYQCDLWSIGCILFELCTGEAL 291  
 CC 341 FQTHDSKEHLAMMERILGPIPOHIQTKR--RKVFHN-QLUMDSSAGRYVR--RRC 395  
 CC 292 FQTHDNEHLAMMERALGPLDEHTRKASGAEKRYRRCGRUMPGANRSRISRAVKRL 351  
 CC 396 KPLKFMILCH-DEHEKFLPLVRMLDEYDPTORTIDELALQHPFF 439  
 CC 352 DRLKDWVSKHVDWTRSRFDLLYGLLAYDPSERLTANEAIDHDFE 396  
 CC  
 CC RESULT 13  
 CC LKHL SCHPO STANDARD; PRT; 575 AA.  
 CC ID LKHL SCHPO  
 CC AC 010156; O9C39; (Created)  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Protein kinase lkh1 (EC 2.7.1.37) (EC 2.7.1.112).  
 CC GN LKHL OR SPACID4.11C.  
 CC OS Schizosaccharomyces pombe (fission yeast).  
 CC CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC CC Schizosaccharomycetes.  
 CC CC NCBI\_TaxID=4896;  
 CC [1]



RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,  
 RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,  
 RA Portetle D., Purnell B., Rechmann S., Rieger M., Rinke M., Rose M.,  
 RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,  
 RA Scheraga A.P., Uristarazu L.A., Vandenbol M., Verhaaselt P.,  
 RA Vierendeels P., Voet M., Volckaert G., Voss H., Wambut R., Wedler E.,  
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hobeisel J.D.;  
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";  
 RA Nature 387:87-90(1997).  
 RL [4]  
 RN  
 RP FUNCTION.  
 RX MEDLINE=97066903; PubMed=910305;  
 RA Lee K., Du C., Horn M., Rabinow L.;  
 RA "Activity and autophosphorylation of LAMMER protein kinases.";  
 RL J. Biol. Chem. 271:27299-27303(1996).  
 RL [5]  
 RN  
 RP AUTOPHOSPHORYLATION.  
 RX MEDLINE=21616999; PubMed=11741326;  
 RA Kim K.-H., Cho Y.-M., Kang W.-H., Kim J.-H., Byun K.-H., Park Y.-D.,  
 RA Bae K.-S., Park H.-M.;  
 RA "Negative regulation of filamentous growth and flocculation by Lkh1,  
 RA a fission yeast LAMMER kinase homolog.";  
 RL Biochem. Biophys. Res. Commun. 289:1237-1242(2001).  
 CC  
 CC -1- BIOCHEM. NONESSENTIAL PROTEIN KINASE.  
 CC  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC  
 CC -1- PTH: Autophosphorylates on all three types of residue.  
 CC  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC  
 CC LAMMER SUBFAMILY.  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>,  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC  
 DR EMBL; M85200; AAA34724.1; -;  
 DR EMBL; 273125; CA97468.1; -;  
 DR EMBL; X97560; CA66171.1; -;  
 DR EMBL; 273123; CA97465.1; -;  
 DR PIR; S64767; S64767;  
 DR SGD; S0003942; KNS1.  
 DR InterPro; IPR000719; Prot. kinase.  
 DR InterPro; IPR002290; Ser./Thr. kinase.  
 DR Pfam; PF00069; Kinase; 1.  
 DR ProDom; PD000001; Prot. kinase; 2.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 KM Transferrase; Serine/threonine-protein kinase; ATP-binding;  
 KM Tyrosine-protein kinase; Phosphorylation.  
 FT DOMAIN 143 146 POLY-SER.  
 FT DOMAIN 148 152 POLY-ASN.  
 FT DOMAIN 313 720 PROTEIN KINASE.  
 FT NP BIND 319 327 ATP (BY SIMILARITY).  
 FT BINDING 343 343 ATP (BY SIMILARITY).  
 FT ACT SITE 440 440 BY SIMILARITY.  
 FT DOMAIN 673 679 SER/THR-RICH.  
 FT CONFLICT 411 421 RFPSSHQAIA -> GSPALISGHC (IN REF. 1).  
 SQ SEQUENCE 737 AA; 83842 MW; F0224445BDE8431 CRC64;  
 Query Match 29.8%; Score 718.5; DB 1; Length 737;  
 Best Local Similarity 33.3%; Pred. No. 1.2e-38;  
 Matches 156; Conservative 90; Mismatches 123; Indels 99; Gaps 11;  
 Oy 67 KGSVRSRRSPKRRNRHNCSSHQSRKSRKRSRTIEDDESHLTCQSGDVLR--RYE 124  
 Db 261 QSNLKQIGSLRKFRTSGSSESAS-----NKSNFKTDKGGHYVQENDIDGSGGRV 314

Oy 125 IVDTLGSGARQKVECTIDHGMQGMHVKYVNVNGVRRAARSEIOVLEHNTDPSNVF 184  
 Db 315 VKDILGQCFKVKYKLCIDNKYEPYVAVKVIKVRADYREAKTELRLQTLNNDDQGF 374  
 Oy 185 RCVOAMLEFHDHGVACIVFELLGLSTYDFKENSFLPFQDHRQMAVQICQSFNLHN 244  
 Db 375 QCLLRREFDVKHNCILVTLDYGRSIVDFMCSNGIARFPSSHQAIAQLRISVCFHDL 434  
 Oy 245 KLHTDLPENILVVKSDVY-----VKNSMKRD-----ERTKATDIKVPDGA 291  
 Db 435 GIHTDLKPENILCDETHIAQKLPKTVOSLSKRREARSKRKIKNPEIKIIFGSA 494  
 Oy 292 TYDEHSHSTVSTHYRAPAEVIALGWSOPDWSICILIEYVLTQVQTHSKHLA 351  
 Db 495 IFHYEHPVSTHYRAPAEVIALGWSOPDWSICILIEYVLTQVQTHSKHLA 554  
 Oy 352 MWERILG-PIPOHNIQKTRK-----KYFHNH--QLDWEHSSAG-- 388  
 Db 555 MMQRLNGTPPTDILDKMFYKSKHKLGNSPSDLNSTVLIKHFDRKTLSLQWPEKRGDTI 614  
 Oy 389 -----RYVRRRC-----KPLKEPMLCH----- 405  
 Db 615 TTESMKRVLOSCLRDIYISKVLDQYGSLSINNNLPPEKWSLSKLAMKRGTHSS 674  
 Oy 406 -----DEBEKLF-----DLYRMLEYPDPTQRTIDEALQHPFDL 441  
 Db 675 SSSSTDDEKKEPFLFWYFIDLLKKEPFPDKRTAKDLDEHMFVL 722  
 RESULT 15  
 DYR2\_HUMAN  
 ID DYR2\_HUMAN STANDARD; PRT; 528 AA.  
 AC Q92630;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Dual-specificity tyrosine-phosphorylation regulated kinase 2  
 DE (EC 2.7.1.1).  
 GN DYRK2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98421512; PubMed=9748265;  
 RA Becker W., Weber Y., Wetzel K., Bimblet K., Tejedor F.J.,  
 RA Joost H.-G.;  
 RT "Sequence characteristics, subcellular localization, and substrate  
 RT specificity of DYRK-related kinases, a novel family of dual  
 RT specificity protein kinases.";  
 RL J. Biol. Chem. 273:25893-25902(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist L.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stadleren M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosk S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunnarsson P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Heltan B., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Rodriguez A.C., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Butlerfield Y.S.N., Krzyzanski M.I., Skalek U., Smalhus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.:  
RT "Generation and initial analysis of more than 15,000 full-length  
RL human and mouse cDNA sequences.";  
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RP SEQUENCE OF 320-528 FROM N.A.

RC TISSUE=Placenta;  
RA Becker W., Voet H.-G.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: IN VITRO, CAN PHOSPHORYLATE HISTONES H3 AND H2B ON SER  
CC AND THR RESIDUES. MAY BE INVOLVED IN THE REGULATION OF CELLULAR  
CC GROWTH AND/OR DEVELOPMENT.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- PTM: AUTOPHOSPHORYLATED ON TYR RESIDUES.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC KMB/DYRK SUBFAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: Y13493; CAA73885.1; -  
DR EMBL: BC005809; AA05809.1; -  
DR EMBL: Y09216; CAA70418.1; -  
DR HSSP: 000534; 1B18.  
DR Genew; HGNC:3093; DYRK2.  
DR MLM; 603496; -  
DR GO; GO:0005737; C:cytoplasm; TAS.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; TAS.  
DR InterPro: IPR000719; Prot Kinase.  
DR InterPro: IPR002290; Ser Thr\_kinase.  
DR Pfam: PF00069; kinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KM Transferrase; Serine/threonine-protein kinase; Tyrosine-protein kinase;  
KW ATP-binding; phosphorylation. PROTEIN KINASE.  
FT DOMAIN 149 462  
FT NP\_BIND 155 163 ATP (BY SIMILARITY).  
FT BINDING 178 178 ATP (BY SIMILARITY).  
FT ACT\_SITE 275 275 BY SIMILARITY.  
SQ SEQUENCE 528 AA; 59714 MW; AF2C6822ED9522D7 CRC64;

Query Match 22.8%; Score 549.5; DB 1; Length 528;  
Best Local Similarity 34.3%; Pred. No. 4.5e-28;  
Matches 121; Conservative 69; Mismatches 122; Indels 41; Gaps 7;

QY 105 DDEGHLICQSGDVLRLRYEIVDTLGEAFGKVCIDHGMDSMVAIVKXNGRYREA 164  
DB 131 DDQGSYVGVPHHVAHYREVLKVIKGSFGQVVKADHKAH-QHVALMVENEKRFHQ 189  
QY 165 ARGEIQLVLEHNSLDNSVFRQVLEWFDHGHVCIIVFELLGLSTYDFIKENSFLPQI 224  
DB 190 AAEELIRLEHRLKQDKONTMNVIMLENTFRNHCMTFELLSDNYELIKKKKFGPSL 249  
QY 225 DAIKQNAVYOICQSLNFLHNKLTHTDLPKENILFVKSIVVVKYNSMKRDERTLKNTDIK 284  
DB 250 PLVKKFRAHSHLQCLDALHKRRIHCDLKPENIL-----LKKQGR-----SGIK 292  
QY 285 VVDFGSATYDDDEHNSLTVSTRHARAEVILALGMSQPCDWSIGCLILEYVIGTFVQTH 344  
DB 293 VDFGSSCYEHQVYVYIQRFRAPREVILGARVGMIDMWSLGCILAEELLTGYPPLPGE 352  
QY 345 DSEKHLAMMERILGPIQHMIOKTRKRYFHHNQ-----LDWDEHSSAGRYVRR 393  
DB 353 DEEDQLACMIETELGMSQKLLDASKRAKXVFSKGVPRYCTVTTLSDGSVVLNGRSRSG 412

QY 394 RCKPLKE-----FMICHDEHEKLF-DLVRRMLEYDPQRTTLDLALONRF 438  
DB 413 KLRGPPESREWGNAKGCDD---PLFLDFLKQCLEWDPXVMTPOQALRHPW 461

Search completed: November 14, 2003, 15:17:48  
Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 15:15:04 ; Search time 40 Seconds

(without alignments)  
2870.835 Million cell updates/sec

Title: US-10-339-656-2

Perfect score: 2410  
Sequence: 1 MCIPLASHSVBBDTHPSHY.....QRITLDEALQHPFDLKKK 445

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2315.5	96.1	481	4	Q9HAZ1
2	1893	78.5	484	4	Q8NSV8
3	1879	78.0	484	6	Q8BGC2
4	1803	74.8	453	4	Q2NRU6
5	1607	66.7	301	11	Q99JL06
6	1604	66.6	301	11	Q8CEU9
7	1437	59.6	498	11	Q91VR2
8	1305	54.1	490	11	Q8CIV1
9	1252	52.0	341	4	Q9BRG8
10	1222	50.7	511	5	Q8IMM0
11	1222	50.7	580	5	Q95RC9
12	1180	49.0	832	5	Q8T041
13	965.5	40.1	903	5	Q8T041
14	963	40.0	409	5	Q8M1Y2
15	839.5	34.8	219	11	Q35721
16	839	34.8	431	10	Q49967

17	802	33.3	434	10	Q943M7	Q943M7 oryza sativ
18	789	32.7	437	10	Q9M598	Q9M598 mesembryant
19	778	32.3	430	10	Q94UC8	Q94UC8 oryza sativ
20	648.5	26.9	881	5	Q9BKX8	Q9BKX8 plasmodium
21	648.5	26.9	881	5	Q8IL19	Q8IL19 plasmodium
22	598	24.8	1300	3	Q8X0V5	Q8X0V5 neurospora
23	550.5	22.8	538	11	Q8BM34	Q8BM34 mus musculus
24	550.5	22.8	586	11	Q922X0	Q922X0 mus musculus
25	549.5	22.8	601	4	Q9BRB5	Q9BRB5 homo sapien
26	546.5	22.7	474	4	Q8WS54	Q8WS54 oikopleura
27	544.5	22.6	568	4	Q9HBV6	Q9HBV6 homo sapien
28	544.5	22.6	588	4	Q9HBV7	Q9HBV7 homo sapien
29	532.5	22.1	722	5	Q9V3D5	Q9V3D5 drosophila
30	524	21.7	722	11	Q8B155	Q8B155 mus musculus
31	523	21.7	508	5	Q20604	Q20604 caenorhabdi
32	523	21.7	796	5	Q9VTF4	Q9VTF4 caenorhabdi
33	523	21.7	817	5	Q9XTF3	Q9XTF3 caenorhabdi
34	522	21.7	461	5	Q8T2V5	Q8T2V5 trypanosoma
35	517	21.5	520	4	Q8NEF2	Q8NEF2 homo sapien
36	509	21.1	607	5	Q917R8	Q917R8 drosophila
37	509	21.1	908	5	Q9VX07	Q9VX07 drosophila
38	495.5	20.6	828	5	Q8MT39	Q8MT39 drosophila
39	483.5	20.1	629	11	Q9EPW2	Q9EPW2 mus musculus
40	480	19.9	608	5	Q8WQ8	Q8WQ8 entamoeba h
41	477	19.8	1157	10	Q9FX39	Q9FX39 arabidopsis
42	474	19.7	882	5	Q8WQ8	Q8WQ8 caenorhabdi
43	474	19.7	948	5	Q22155	Q22155 caenorhabdi
44	472	19.6	1138	10	Q8RX85	Q8RX85 arabidopsis
45	472	19.6	1266	10	Q9LSH3	Q9LSH3 arabidopsis

## ALIGNMENTS

## RESULT 1

Q9HAZ1 PRELIMINARY: PRT: 481 AA.

AC Q9HAZ1, 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Protein serine/threonine kinase CLK4.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21100912; PubMed=11170754;  
RA Schultz J., Jones T., Bork P., Sheer D., Blencke S., Steyrer S.,  
RT Wellbrock U., Beyec D., Ullrich A., Wallasch C.,  
RT "Molecular Characterization of a cDNA Encoding Functional Human CLK4  
RT Kinase and Localization to Chromosome 4q35."  
RL Genomics 71:368-370(2001).  
RN (2)  
RP SEQUENCE FROM N.A.  
RA Schultz J., Jones T., Bork P., Sheer D., Blencke S., Steyrer S.,  
RT Wellbrock U., Beyec D., Ullrich A., Wallasch C.,  
CC Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AF294429; AAG10074.1; --  
DR HSSP; Q00534; 1B18.  
DR Gene; HGNC:13659; CLK4.  
DR InterPro: IPR000719; Prot kinase.  
DR InterPro: IPR002290; Ser. Thr. kinase.  
DR Pfam; PF00069; Pkinase; 1.  
DR ProDom; PD000001; Prot. kinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR PROSITE; PS00107; PROTEIN KINASE\_ATP; 1.  
DR PROSITE; PS00101; PROTEIN KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 481 AA; 57492 MW; F402C3685CDA306 CRC64;

Query Match 96.1%; Score 2315.5; DB 4; Length 481;  
Best Local Similarity 96.2%; Pred. No. 1,9e-186;  
Matches 432; Conservative 2; Mismatches 4; Indels 11; Gaps 2;

QY 8 SH-SVEEDT-----PSHYLEARSINERDVRVDEYNDYCEGYPRRYHND 56  
DB 33 SHSSTQENRCKKHQKESDCHYLEARSINERDVRVDEYNDYCEGYPRRYHND 92  
QY 57 IEQYRIHCKSSVSRSSPKRKRNHCKSSHOSRSKSHRKRKRSIEDDEGHILCOG 116  
DB 93 IEQYRIHCKSSVSRSSPKRKRNHCKSSHOSRSKSHRKRKRSIEDDEGHILCOG 152  
QY 117 DVLRAREIVDTLGEAGFGVVEICDHGMGMHVAKIYVNGRYEARSEIQTVEHIN 176  
DB 153 DVLRAREIVDTLGEAGFGVVEICDHGMGMHVAKIYVNGRYEARSEIQTVEHIN 212  
QY 177 STDPNSVFCVOMLEWFDHGHVCIYFELLGLSTYDFIKENSLFPQIDHIRMAVOICQ 236  
DB 213 STDPNSVFCVOMLEWFDHGHVCIYFELLGLSTYDFIKENSLFPQIDHIRMAVOICQ 272  
QY 237 SINPLHNKLTHTDLKPENLTFYKSDYVYVYKSKMRDERTLKNTDIXVDFGSAITYDE 296  
DB 273 SINPLHNKLTHTDLKPENLTFYKSDYVYVYKSKMRDERTLKNTDIXVDFGSAITYDE 332  
QY 297 HNSTLVSSTRYRAPEVILALGMSQPCDWSIGCILLEYVIGFTVPFOTHSKEHLAMMERI 356  
DB 333 HNSTLVSSTRYRAPEVILALGMSQPCDWSIGCILLEYVIGFTVPFOTHSKEHLAMMERI 392  
QY 357 LGPIPOHMOKTRKRYFHNNOLDMDEHSAGRYVRRCKPLKEFMLCHDEHEKLFDLV 416  
DB 393 LGPIPOHMOKTRKRYFHNNOLDMDEHSAGRYVRRCKPLKEFMLCHDEHEKLFDLV 452  
QY 417 RMLEVDPTQRTITLDEALQHPFDLLKK 445  
DB 453 RMLEVDPTQRTITLDEALQHPFDLLKK 481

RESULT 2

Q8NSV8 PRELIMINARY; PRT; 484 AA.  
AC Q8NSV8;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE CMC-like kinase 1.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=bone;  
RA Strausberg R.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC031549; AA31549.1; -  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.  
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
KW ATP-binding; Kinase; Transferase.  
SQ SEQUENCE 484 AA; 57290 MW; F34B5B44988BD118 CRC64;

Query Match 78.5%; Score 1893; DB 4; Length 484;  
Best Local Similarity 82.3%; Pred. No. 6.8e-151;  
Matches 353; Conservative 30; Mismatches 44; Indels 2; Gaps 2;

QY 18 SHYLEARSINERDVRVDEYNDYCEGYPRRYHNDYCEGYPRRYHNDYCEGYPRRYHND 77  
DB 54 SHYLEARSINERDVRVDEYNDYCEGYPRRYHNDYCEGYPRRYHNDYCEGYPRRYHND 113  
QY 78 KRK-RNRHCKSSHOSRSKSHRKRKRSIEDDEGHILCOGSDVLRAREIVDTLGEAGFG 135  
DB 114 KSKRIHNSHTRSHOSRSKSHRKRKRSIEDDEGHILCOGSDVLRAREIVDTLGEAGFG 173  
QY 136 KVEECIDHGMGMHVAKIYVNGRYEARSEIQTVEHINSTDPSVFCVOMLEWFDH 195  
DB 174 KVEECIDHGMGMHVAKIYVNGRYEARSEIQTVEHINSTDPSVFCVOMLEWFDH 233  
QY 196 HGVCIVFELLGLSTYDFIKENSLFPQIDHIRMAVOICQINFLHNKLTHTDLKPE 255  
DB 234 HGVCIVFELLGLSTYDFIKENSLFPQIDHIRMAVOICQINFLHNKLTHTDLKPE 293  
QY 256 ILFVQSDYVYVYKSKMRDERTLKNTDIXVDFGSAITYDEHNSTLVSSTRYRAPEVILA 315  
DB 294 ILFVQSDYVYVYKSKMRDERTLKNTDIXVDFGSAITYDEHNSTLVSSTRYRAPEVILA 353  
QY 316 LGMSQPCDWSIGCILLEYVIGFTVPFOTHSKEHLAMMERILGPIPOHMOKTRKRYFH 375  
DB 354 LGMSQPCDWSIGCILLEYVIGFTVPFOTHSKEHLAMMERILGPIPOHMOKTRKRYFH 413  
QY 376 HNOLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLFDLVRMLEVDPTQRTITLDEALQ 435  
DB 414 HDRLDMDHSSAGRYVRRCKPLKEFMLCHDEHEKLFDLVRMLEVDPTQRTITLDEALQ 473  
QY 436 HPPFDLLKK 444  
DB 474 HPPFDLLKK 482

RESULT 3

Q9BGQ1 PRELIMINARY; PRT; 484 AA.  
AC Q9BGQ1;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Hypothetical 57.3 kDa protein.  
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9541;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Frontal cortex;  
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
SUzuki Y., Sugano S., Hashimoto K.;  
RT "Isolation of full-length cDNA clones from macaque brain cDNA  
libraries."  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AB056421; BAB33079.1; -  
DR HSSP; O00534; 1818  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.  
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.  
KW Hypothetical protein; ATP-binding; Kinase;  
KW Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 484 AA; 57318 MW; F5324D4007119878 CRC64;

Query Match 78.0%; Score 1879; DB 6; Length 484;  
Best Local Similarity 79.3%; Pred. No. 1e-149;  
Matches 356; Conservative 30; Mismatches 51; Indels 12; Gaps 4;



```

QY 8 SH$VEED-----TH-----PSHYLARSJNEDDYDRRYVEYENDYCEGVPRPHYRD1 57
Db 34 SH$S$AENRCKRKYTH$KMD$SHYLSRCLNE$DYHSRKYIDYERNDYNQCEPRH$RDH 93
QY 58 ESGYRHC$K$SVRSR$SR$PKR-RNRHCSH-Q$R$K$H$R$K$R$S$I$EDDEB$H1CQ$ 115
Db 94 ESRYONH$S$K$S$GR$SR$S$Y$K$K$H$R1H$H$T$H$R$H$G$H$R$K$R$K$R$T$R$VEDEB$E$H1CQ$ 153
QY 116 G$V$L$R$R$Y$E1V$D$L$G$A$F$G$K$V$E$C1D$H$O$M$H$A$V1Y$K$N$O$R$Y$E$A$R$S$E1Q$Y1E$H1 175
Db 154 G$V$L$S$R$Y$E1V$D$L$G$A$F$G$K$V$E$C1D$H$A$G$R$H$A$V1Y$K$N$O$R$Y$E$A$R$S$E1Q$Y1E$H1 213
QY 176 N$T$D$P$A$S$V$F$C$V$O$M$L$E$M$F$D$H$G$H$V$C1V$E$P$E$L1G1$T$Y$D$E1K$E$N$F$P$O1D$H1R$O$M$A$Y$O1C 235
Db 214 N$T$D$P$M$T$R$C$V$O$M$L$E$M$F$E$H$G$H1C1V$E$P$E$L1G1$S$T$Y$D1K$E$N$G$F$P$R$D$H1R$O$M$A$Y$O1C 273
QY 236 Q$S1N$F$L$H$N$K$L$T$H$D1K$E$N1L$F$V$K$D$Y$V$V$K$T$N$K$M$K$D$E$R$T$L$N$D1K1V$D$F$G$A$T$Y$D 295
Db 274 K$V$N$F$L$H$S$N$K$L$T$H$D1K$E$N1L$F$V$O$D$Y$E$A$N$P$K1K$E$D$E$R$T$L1N$D1K1V$D$F$G$A$T$Y$D 333
QY 296 E$H$S$T1V$N$R$H$A$P$R$V$U1L$G$M$O$P$C$D$W$S1G$C$L1E$Y$G1F$V$F$O$H$S$K$E$L$A$M$R 355
Db 334 E$H$S$T1V$N$R$H$A$P$R$V$U1L$G$M$O$P$C$D$W$S1G$C$L1E$Y$G1F$V$F$O$H$S$K$E$L$A$M$R 393
QY 356 I$L$P1E$O$H1O$K$T$R$K$E$F$H$N$Q$L$D$M$D$E$H$S$A$G$R$Y$V$R$C$K$P$L$E$F$M$L$C$H$D$E$B$E$K$L$F$D$L 415
Db 394 I$L$P$L$P$E$H$M1O$K$T$R$K$E$F$H$N$R$L$D$M$D$E$H$S$A$G$R$Y$V$R$C$K$P$L$E$F$M$L$S$Q$V$E$H$C$L$F$D$L 453
QY 416 V$R$M$L$E$Y$D$P$O$R1T$L$D$E$A$L$O$H$E$P$F$D$L$K 444
Db 454 I$Q$K$E$L$D$E$P$A$K$R1T$L$K$E$A$L$H$P$F$D$L$K 482

```

Q9NRL6	PRELIMINARY;	PRT;	453 AA.
ID	Q9NRL6		
AC	Q9NRL6;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	CLK4.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	1)		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Bone marrow;		
RA	Zhao M., Song H., Li N., Peng Y., Han Z., Chen Z.;		
RT	"A novel gene expressed in human bone marrow."		
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBS databases.		
CC	-!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.		
DR	EMBL; AF12224; AAB87326.1; -.		
DR	HSSP; 000534; 1B18.		
DR	InterPro; IPR000719; Prot_kinase.		
DR	InterPro; IPR002290; Ser_Thr_pKinase.		
DR	Pfam; PF00069; pkinase; 1.		
DR	ProDom; PD000001; Prot_kinase; 1.		
DR	SMART; SM00220; S_TKc; 1.		
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.		
DR	PROSITE; PSS0011; PROTEIN KINASE DOM; 1.		
DR	PROSITE; PS00108; PROTEIN_KINASE_SF; 1.		
DR	ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.		
QO	SEQUENCE 453 AA; 53806 MW; 3B2AC6A3FFAE9461 CRC64;		

Query Match 74.8%; Score 1803; DB 4; Length 455;  
Best Local Similarity 80.4%; Pval No. 2,4e-13;  
Matches 344; Conservative 16; Mismatches 40; Indels 28; Gaps 3

Ox 18 SHLRLKSLNKKDYRDRVDEYNDYCEGYPPPHYHDIISGTRICKSSSVTSRRSP 77  
|||||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||  
Db 54 SHLLERSLNKKDYSRRIIDRYNDYQGGEPGRPDHSRRQWMSKSSSGRSRSSY 113  
|||||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||S|||

Qy	78	KRKRNNHSGHSSQSRKSRKRRKRSRAIEDDEBGHLICSGDVLRYARYEIVYPTLLEGAGKY	137
Db	114	KSKGRH-----HSTSRSH-----GD-----EIVPTLLEGAGKY	145
Qy	138	VEECIDHGDGMHVAVKIVAGVGRYREARSEIÖVLEHLNSTDPNSVRCVÖMLEMFHDHG	197
Db	146	VECIDHKGAGRRVAVKIVANDVRYCEARSEIÖVLEHLNNTDPSRSTRCVÖMLEMFHDHG	205
Qy	198	HVCIVELLIGLSTYFIKENSFLPQIDHICQAYÖICQSNFNLHNKLTHYDLKPBENIL	257
Db	206	HICIVELLIGLSTYFIKENGLPRLDHIKRMAYÖICQSNFNSKMLTHYDLKPBENIL	265
Qy	258	FVSDVYVYKXNKMDEBRTLKNPDIKVYDGSATYDDEHSTLVSRHRAPRVPILALG	317
Db	266	FVSDYTEALNPKIKDERTLINPDIKVYDGSATYDDEHSTLVSTRHRAPRVPILALG	325
Qy	318	WSQPCVWSMIGCLILEYVGFVFOGTHSKHLLAMERILPPIQOHMIÖKTRKKEYPHN	377
Db	326	WSQPCVWSMIGGLINEYIGFVFOGTHSKHLLAMERILPPIQOHMIÖKTRKKEYPHN	385
Qy	378	ÖLDWDEHSSAGRYVRRCPLKEFMLCHDEHEKLFDIVRMLEYDPTÖRTTDEALÖHP	437
Db	386	ÖLDWDEHSSAGRYVRRCPLKEFMLCHDEHEKLFDIVRMLEYDPTÖRTTDEALÖHP	445
Qy	438	FFDLKKK 445	
Db	446	FFDLKKK 453	

RESULT 5			
099UT6			
ID	099UT6	PRELIMINARY;	PRT; 301 AA.
AC	Q99JL6		
DT	01-JUN-2001	(TREMBLrel. 17, Created)	
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)	
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)	
DE	Similar to CDC like kinase 4.		
GN	CLK4.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Strausberg R.;		
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.		
CC	-i- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.		
DR	EMBL; BC002220; AA002220.1; -.		
DR	MGJ; MG1096551; CLK4.		
DR	InterPro; IPR000719; Prot_kinase.		
DR	InterPro; IPR002290; Ser_thr_kinase.		
DR	Pfam; PF00069; Pkinase; 1.		
DR	ProDom; PS00001; Prot_kinase; 1.		
DR	PROSITE; PSS00011; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.		
KW	ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.		
SO	SEQUENCE 301 AA; 35768 MW; 761757E2E28C6C1 CRC64;		

Query Match	66.7%	Score 1607	DB 11	Length 301
Best Local Similarity	98.3%	Pred. No. 4.4e-127		
Matches 296	Conservative 3	Mismatches 12	Indels 0	Gaps 0
QY	145	MDGNHVAVKIYKNGRYEARARSEIQVLEHINSTDPNSVFCRCOMLEWFDHGHVCVFE	204	
Db	1	MDGHVAVKIYKNGRYEARARSEIQVLEHINSTDPNSVFCRCOMLEWFDHGHVCVFE	60	
QY	205	LLGLSTVDFIKENSLFPQIDHIROMAYQICQSIINFLHHNKLTTTDLKPENILTVKSDYV	264	
Db	61	LLGLSTVDFIKENSLFPQIDHIROMAYQICQSIINFLHHNKLTTTDLKPEITILFVKSQYV	120	
QY	265	VKYNSSKKRDBERTIKATNDIKVYDEGSATYUDEHNSHSTVSTRHYAPVETIALGMSQPCDV	324	
Db	121	VKYNSSKKRDBERTIKATNDIKVYDEGSATYUDEHNSHSTVSTRHYAPVETIALGMSQPCDV	180	



QY 325 WSICGILIEYLGFTVQTHDSKHEMLAMERILGPIDHMIQKTRKRYPHHNOJLWDEH 384  
 DB 181 WSICGILIEYLGFTVQTHDSKHEMLAMERILGPIDHMIQKTRKRYPHHNOJLWDEH 240  
 QY 385 SSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLDEALQHPFDLLKX 444  
 DB 241 SSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLDEALQHPFDLLKX 300  
 QY 445 K 445  
 DB 301 K 301

## RESULT 6

08CEU9 PRELIMINARY; PRT; 301 AA.  
 AC 08CEU9;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE CDC like kinase 4.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RA MEDLINE=22354683; PubMed=12466851;  
 RA THE FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RL 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL: AK013974; BAC25420.1;  
 SQ SEQUENCE 301 AA; 35740 MW; 58EADD0DD228C6DE CRC64;

Query Match 66.6%; Score 1604; DB 11; Length 301;  
 Best Local Similarity 98.0%; Pred. No. 7.8e-127;  
 Matches 295; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 145 MDGHWAVKTVKNGRYREARASEIQLYELHNSDPNSVRCVOMLEWPFHGHVCIIVE 204  
 DB 1 MDGHWAVKTVKNGRYREARASEIQLYELHNSDPNSVRCVOMLEWPFHGHVCIIVE 60  
 QY 205 LAGSTYDFIKENSFLPFOIDHIRQAVYQICQINFPHNKLTHTDLKPENILFVKSDDY 264  
 DB 61 LAGSTYDFIKENSFLPFOIDHIRQAVYQICQINFPHNKLTHTDLKPENILFVKSDDY 120  
 QY 265 VKVNSKMRDERTLKNTDIKVVDGSAFYDDEHSTLSTRHRYRAPVILALGMSQPCV 324  
 DB 121 VKVNSKMRDERTLKNTDIKVVDGSAFYDDEHSTLSTRHRYRAPVILALGMSQPCV 180  
 QY 325 WSICGILIEYLGFTVQTHDSKHEMLAMERILGPIDHMIQKTRKRYPHHNOJLWDEH 384  
 DB 181 WSICGILIEYLGFTVQTHDSKHEMLAMERILGPIDHMIQKTRKRYPHHNOJLWDEH 240  
 QY 385 SSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLDEALQHPFDLLKX 444  
 DB 241 SSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLDEALQHPFDLLKX 300  
 QY 445 K 445  
 DB 301 K 301

## RESULT 7

QY 091YR2 PRELIMINARY; PRT; 498 AA.  
 AC 091YR2;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE CDC-like kinase 2.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Strubberg R.,  
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC015080; AAH15080.1;  
 DR InterPro: IPR001064; Crystalin.  
 DR InterPro: IPR000719; Prot. kinase.  
 DR InterPro: IPR02290; Ser. thr. kinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; Prot. kinase; 1.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR PROSITE: PS00225; CRYSTALLIN BETA-DOMAIN; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00101; PROTEIN KINASE DOW; 1.  
 DR PROSITE: PS00108; PROTEIN KINASE ST; 1.  
 KW ATP-binding; kinase; Transferase.  
 SQ SEQUENCE 498 AA; 59956 MW; 0A87B74D78869B19 CRC64;

Query Match 59.6%; Score 1437; DB 11; Length 498;  
 Best Local Similarity 60.6%; Pred. No. 1.7e-112;  
 Matches 262; Conservative 69; Mismatches 95; Indels 6; Gaps 3;

QY 18 SHYLEARL---NRRDVRDRRYDEY-RNRYCEGYVRRHHRIDESGRHCKSSVRS 72  
 DB 50 SHYRSRYDHSRRYDRYCGSYKRNYSRDRGEAYDTDFPROSEYHRENSYSRS 109  
 QY 73 RRS-PRKRNHCSHOSRSHRRKRSIEDDEGHLFCQSGDVRAVEIYDTLGE 131  
 DB 110 QRSSRRKRRRRRRRTSRSSSHSRRAKVEDDABGHLTYHODMLOERYEIVSTIGE 169  
 QY 132 GAFGVVECIDHGMDCGAVAVKTVKNGRYREARASEIQLYELHNSDPNSVRCVOMLE 191  
 DB 170 GTFGRVVOCVDRRSGTVALKIIKVEKYEAARLEINVEKINEKDPENKLCVQWFD 229  
 QY 192 WFDHGHVCIYELIGSTYDFIKENSFLPFOIDHIRQAVYQICQINFPHNKLTHTDL 251  
 DB 230 WFDHGHVCIYELIGSTYDFIKENSFLPFOIDHIRQAVYQICQINFPHNKLTHTDL 289  
 QY 232 KEENILFYKSDYVVKYNSKMRDERTLKNTDIKVVDGSAFYDDEHSTLSTRHRYAPE 311  
 DB 290 KEENILFYNSDYELTYNLEKRDERSVSTAVRVVDGSAFYDDEHSTLSTRHRYAPE 349  
 QY 312 VILAGMSQPCDWSIGCILEYLGFTVQTHDSKHEMLAMERILGPIDHMIQKTRK 371  
 DB 330 VILAGMSQPCDWSIGCILEYLGFTVQTHDSKHEMLAMERILGPIDHMIQKTRK 409  
 QY 372 KYFHNOJLWDEHSSAGRYVRRCKPLKEFMLCHDEHEKLPDLVRMLEYDPTQRIITLD 431  
 DB 410 KYFYNGRLDWBENTAGRYVRENCPLARRYITLSEADHGFJDIENNLLEYEPKRLTLG 469  
 QY 432 EALQHPFDLLKX 443  
 DB 470 EALQHPFDLLKX 461

## RESULT 8

QY 08CIV1 PRELIMINARY; PRT; 490 AA.  
 AC 08CIV1;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE CDC-like kinase 3.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team,  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK090215; BAC41138.1;  
 SQ SEQUENCE 490 AA; 58469 MW; 2B2B71B9A3DA1D2 CRC64;  
  
 Query Match 54.1%; Score 1305; DB 11; Length 490;  
 Best Local Similarity 55.4%; Pred. No. 2.1e-101;  
 Matches 247; Conservative 62; Mismatches 111; Indels 26; Gaps 4;  
  
 QY 4 PLEASHVEDTHPSHYLAKSLNERDYDRDYVDYKNDYCEGYPRHYHDISGRI 63  
 DB 43 PPRSRSRSHDRIP-----YQRYRHRDSDTYR--CEERSPSFG---EDCYGS 86  
 QY 64 HCSKSVSRSSRSPKR-----KRNHCSSHQSRKSHRKRKRSIEDDEGHLIC 113  
 DB 87 SRSHRRSRREPAPRYRKRKAHGHKRRTRSCSSASRSQSSKRSRVEDDKGHLVC 146  
 QY 114 QSGDVLARVEIYDTLGEAPFGYVVCIDHGMGMVAVKYKNGRYEARSEI QYLE 173  
 DB 147 RIGDWLOERYEIVGNIGEGTFGKVCCLDHARGSKVALKIRNVGKYEARLEINVLK 206  
 QY 174 HLNSTDNSVFCVOMLEMFHGHVCIYFELLGLSTYDFIKENSFLPFOIDHROMAYQ 233  
 DB 207 KIKKDKENKFLCVMSDMWNFHGMCIAPFELLGKTPFELENNQPPPLPHVHMAVQ 266  
 QY 234 IQCSINFLHNKLTHTDLKPNILFYKSDYVVKYNSKMRDELTNDIKYVDFGSATY 293  
 DB 267 LCHALRPLHNMOLAHDTLKPENILFYVSEFETLYNHSKCEKSVANTISIRVADGSAIF 326  
 QY 294 DDEHSTVSTRYRABEVILALGWSQPCDWSIGCIIIEYLYGFVPGTDSKHLMM 353  
 DB 327 DHEHTITVATRYRPREVILELGMAPCDVMSIGCIIIEYLYGFVPGTDSKHLMM 386  
 QY 354 ERILGPIPMIOKTRKRYFHHNQLDWDEHSAGRYVRRCKPLKEFMLCHDEHEKLF 413  
 DB 387 EKILGPIPSMIRHKRYKQFYKGLVMDENSDRGYVKNCKPLKSYMLQSLHVOVF 446  
 QY 414 DLVARMLEYPDRTITDEALQHPFF 439  
 DB 447 DLVARMLEYPDRTITDEALQHPFF 472  
  
 RESULT 9  
 Q9BRG8 PRELIMINARY; PRT; 341 AA.  
 ID Q9BRG8  
 AC Q9BRG8  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DB Similar to CDC-like kinase 2 (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovary;  
 RA Strauberg R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBS databases.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR EMBL; BC006374; AA06274.1;  
 DR InterPro; IPR001064; Crystalin.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Prodom; PD000001; Prot\_kinase; 1.

DR SMART: SM00220; 5\_TKC; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 FT NON TER 1  
 SQ SEQUENCE 341 AA; 40236 MW; AF43DA2455A5AD4 CRC64;  
  
 Query Match 52.0%; Score 1252; DB 4; Length 341;  
 Best Local Similarity 67.7%; Pred. No. 3.9e-97;  
 Matches 218; Conservative 53; Mismatches 51; Indels 0; Gaps 0;  
  
 QY 122 RYELVDLGGAGKGVCEICDHGMGMVAVKYKNGRYEARSEI QYLEHNSITDPK 181  
 DB 4 RYELVSTLGGRTGERVQCVDRHGGARRVALKIKVVEKKEARLEINLEKINERDP 63  
 QY 182 SVPRCVOMLEMFHGHVCIYFELLGLSTYDFIKENSFLPFOIDHROMAYQIQSINFL 241  
 DB 64 NKNLCVQMPMFVDYHGMCIAPFELLGKTPFELENNQPPPLPHVHMAVQ 123  
 QY 242 HNNKLTHTDLKPNILFYKSDYVVKYNSKMRDELTNDIKYVDFGSATYDEHNSIT 301  
 DB 124 HDNKLTHDTLKPENILFYVSEFETLYNHSKCEKSVANTISIRVADGSAIF 183  
 QY 302 VSTRHYRABEVILALGWSQPCDWSIGCIIIEYLYGFVPGTDSKHLMMERILGPIP 361  
 DB 184 VSTRHYRABEVILALGWSQPCDWSIGCIIIEYLYGFVPGTDSKHLMMERILGPIP 243  
 QY 362 QHMIQTRKRYFHHNQLDWDEHSAGRYVRRCKPLKEFMLCHDEHEKLFVLVARMLE 421  
 DB 244 SRMTRKRYKQFYKGLVMDENSDRGYVKNCKPLRRLYLSAEHHLQFLIESMLE 303  
 QY 422 YDPTORTITDEALQHPFFDLK 443  
 DB 304 YEPARKLTITGEALQHPFFARLR 325  
  
 RESULT 10  
 Q8IMMO PRELIMINARY; PRT; 511 AA.  
 ID Q8IMMO  
 AC Q8IMMO  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE CG1658-PB.  
 GN DOA.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abriil J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Chert J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.R., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dudzon K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Mei M.H., Ibegam C.,  
RA Talati M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,  
RA Kimmel B.E., Kodira C.D., Kratz C., Kravitz S., Kulip D., Lai Z.,  
RA Lasko P., Lei Y., Leitesky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Meriklov G., Miletina N.V., Moberly C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon S., Ruskern D.R., Paele J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reibert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Spidling A.C., Simpson M., Skupski M.P., Smith T.,  
RA Svirskas R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodgett M., Wray K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195 (2000).  
[2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banson J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,  
RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,  
RA Pacieb J., Paragas V., Park S., Patel V., Richards S., Scheeler F.,  
RA Phouanavong S., Pittman G.S., Puri V., Venter E., Venter J.C.,  
RA Stapleton M., Strong S., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.B., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Seattle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.,  
RT "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,  
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
[5]  
RP SEQUENCE FROM N.A.  
RA Flybase;  
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AB003767; AAF5683.2; -  
SQ SEQUENCE 511 AA; 60210 MW; 806366BFBDB8A948 CRC64;

Query Match 50.7%; Score 1222; DB 5; Length 511;  
Best Local Similarity 57.8%; Pred. No. 2.2e-94;  
Matches 230; Conservative 58; Mismatches 84; Indels 26; Gaps 4;

QY SVSRSSSPKRRKRNHCSSHQ---SRSKHRRKRSRS-----IE 104  
DB 86 SKRSRHSKPPASRRQHKRYRDETHSSSRHRDRADDERDSSGRNNRQSAKTAAPVQ 145  
QY DDEBHLTGOSGDVLRARVEYDITGEGAFGVVECDIHGMGMHVAIVKVGARYREA 164  
DB 146 DDADGHLIYHTGDIHHRKYIMATGEGTFGVVVKVMDERD-YCMALKIKNKEXYREA 204  
QY ARSEIQLVLEHNSLTPNSVFRVCVOMLEWFDHGHGVCIVFELLGLSTYDFIKENSFLPOI 224  
DB 165 ARSEIQLVLEHNSLTPNSVFRVCVOMLEWFDHGHGVCIVFELLGLSTYDFIKENSFLPOI 224

DB 205 AKLEINALEKIAQKQDCHCHLCVYKMIWPDYHGHMCIVFEMGLSVFDFLRNNYEPVPL 264  
QY 225 DHIRQNAVYOICOSINFLHNKLTHTDLPKPNILFYKSDIVVYKNSKMRDETLANTDVK 284  
DB 265 DQVRHNAAYOLCVSKVELHNRRLHTHDLKPNENILFYDSIDTSHYXNKHINEVRVAVXTDVR 324  
QY 285 VVDEGSATYDDEHSHVSTHRYRAPEVITLACMSGPCVMSIGCILEFYVGFVFOCH 344  
DB 325 LIDFGSATYDDEHSHVSTHRYRAPEVITLACMSGPCVMSIGCILEFYVGFVFOCH 384  
QY 345 DSKHEHLLAMERLIGPIPOHMIQTRKRYEHHNOLDWDESSAGRYVRRCRPLKEFMVC 404  
DB 385 DNEHLLAMERLIGQIPYMARKT-KTKFYHGKLDMDKSSAGRYVRRCRPLKEFMVC 443  
QY 405 HDEHEKLPDLVRKMLEYPTORTILDEALQHPFDLL 442  
DB 444 DSEDHCELSLJKKMLEYPPSSRITLGEALHHPFDRL 481

RESULT 11  
ID 095RC9 PRELIMINARY; PRT; 580 AA.  
AC 095RC9;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
GN LD44053p.  
GN DOA OR CG1658.  
OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Berkeley;  
RA Stapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Fartan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacieb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,  
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AY061474; AAL29022.1; -  
DR Flybase; FBgn0000480; Doa.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Prodom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKc1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Transferase.  
SQ SEQUENCE 580 AA; 67894 MW; F73AC5B4773AF3A CRC64;

Query Match 50.7%; Score 1222; DB 5; Length 580;  
Best Local Similarity 57.8%; Pred. No. 2.5e-94;  
Matches 230; Conservative 58; Mismatches 84; Indels 26; Gaps 4;

QY SVSRSSSPKRRKRNHCSSHQ---SRSKHRRKRSRS-----IE 104  
DB 155 SKRSRHSKPPASRRQHKRYRDETHSSSRHRDRADDERDSSGRNNRQSAKTAAPVQ 214  
QY 105 DDEBHLTGOSGDVLRARVEYDITGEGAFGVVECDIHGMGMHVAIVKVGARYREA 164  
DB 215 DDADGHLIYHTGDIHHRKYIMATGEGTFGVVVKVMDERD-YCMALKIKNKEXYREA 273  
QY 165 ARSEIQLVLEHNSLTPNSVFRVCVOMLEWFDHGHGVCIVFELLGLSTYDFIKENSFLPOI 224  
DB 274 AKLEINALEKIAQKQDCHCHLCVYKMIWPDYHGHMCIVFEMGLSVFDFLRNNYEPVPL 333  
QY 225 DHIRQNAVYOICOSINFLHNKLTHTDLPKPNILFYKSDIVVYKNSKMRDETLANTDVK 284



```

Db      570 RENNVPFYPDQVRHNAVQLCYSAKFLDHNRLTHLDKPNILFVDSQYTGSHYMKIARE 629
Qy      275 ERLTKRTDVKYVDFGSAATYDDHNSITVSTRYRAPVILALGNSQPCDWSIGCILLEY 334
Db      630 VRRVKNQTVRLIDFGSATFDEHSHSTVSTRYRAPVILELGMQPCDWSIGCILFEL 689
Qy      335 YGCFVYFQTHDSKHEHLLAMMERILGPIF-----QHMIQKTRKKYFHHNQLDDESSAGR 389
Db      690 YIGITLFFQTHDRHLLAMMERILGQIPYMAANHILSKTKTKYFHHKLDMDDESSAGR 749
Qy      390 YVRRRCKPLKEFMLCHDEHEKLFVLRMLEYDPTQRTITLDEALQHPFDLL 442
Db      750 YVRDHCKPLFLCOLSPSEDHCELFLIKKMLEYEPSSRTITGELALHPFFRL 802

```

## RESULT 13

```

ID      Q17917 PRELIMINARY; PRT; 903 AA.
AC      Q17917; Q19042;
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DT      01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      E02H4.3 protein.
GN      E02H4.3.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Sims M.;
RL      Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=94150718; PubMed=7906398;
RA      Wilson R., Almscough R., Anderson K., Baynes C., Berke M., Coulson A.,
RA      Craxton M., Dear S., Du Z., Durbin R., Faveille A., Fulton L.,
RA      Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA      Jones M., Kershaw J., Kistler J., Laister N., Latreille P.,
RA      Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA      Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA      Smailon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA      Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA      Watson A., Weinstock L., Wilkinson-Spoat J., Woldman P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT      elegans."
RL      Nature 368:32-38(1994).
RN      [3]
RP      SEQUENCE FROM N.A.
RA      Barlow K.;

```

```

RL      Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC      -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR      EMBL; Z70205; CAA94122.1; -.
DR      EMBL; Z68003; CAA94122.1; JOINED.
DR      EMBL; Z68003; CAA91979.1; -.
DR      EMBL; Z70205; CAA91979.1; JOINED.
DR      InterPro; IPR000719; Prot kinase.
DR      InterPro; IPR002965; P_tich_extensn.
DR      InterPro; IPR002290; Ser_thr_kinase.
DR      Pfam; PF00069; PKINASE; 1.
DR      PRINTS; PR01217; PRICHKXTNSN.
DR      ProDom; PD000001; Prot kinase; 1.
DR      SMART; SM00220; S_TKc; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR      PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR      ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ      SEQUENCE 903 AA; 98863 MW; 3EB78CE79C9DA187 CRC64;

```

Query Match 40.1%; Score 965.5; DB 5; Length 903;  
Best Local Similarity 46.6%; Pred. No. 1.7e-72;

```

Matches 185; Conservative 72; Mismatches 99; Indels 41; Gaps 6;
Qy      66 SKSASVSRSSPK-----RKRNRHSSQSSKSKRRKRSISIEDDEGHILCOGQDVL 119
Db      507 SRSGLOASQARPPVPEIVSSQRTQ-----QDDKQGHLYSKGDFI 547
Qy      120 RAREYIVDTLGEAGFGKVCIEDHGMGNVAVIKVNGRYEARSEIQVLEHNSRD 179
Db      548 IARFTIYDILGBOGTFGNVR-VNDSLSDTFMAKIKIKNSKYFEAKLEKVLQKLAED 606
Qy      180 PNVFRCQMLENPFDDHGHVCIVFELLGSTYDFIKNSFLPQIDHIKQMAVQCSIN 239
Db      607 PEKKNWIIHNGSYFDYNGHICLFLDMGSGSIFDFLKNYKPYPMEOQTILITQLCNAVK 666
Qy      240 FLHNNKLTHTDLPENILFPKSDYVVKYNSK-----MKRDE-----RTLKNTDIKY 285
Db      667 FLHNNKLTHTDLPENILFVDSRYTKLVYDKKYGKVLARFKIDDFQPLVHSHVRL 726
Qy      286 VDFGSATYDDHNSITVSTRYRAPVILALGNSQPCDWSIGCILLEYLGFVQTHD 345
Db      727 IDFGSATFDEHSHSTVSTRYRAPVILELGMQPCDWSIGCILLEYLTVGTLFQTHE 786
Qy      346 SKHEHLLAMMERILGPIQHMIOKTRKKYFHHNQLDDESSAGRYYRRCKPLKEMLCH 405
Db      787 NREHLAMMERVLDIDPLRVAKRT-KTFPIINGRLDWNTSADAAVVRDCKPLRSMST 845
Qy      406 DEHEKLFVLRMLEYDPTQRTITLDEALQHPFDLL 442
Db      846 DPEHVELFELIEMLMMEPLARKKLPALQHPHFRNL 882

```

## RESULT 14

```

ID      Q08MY2 PRELIMINARY; PRT; 409 AA.
AC      Q08MY2;
DT      01-OCT-2002 (TrEMBLrel. 22, Created)
DT      01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      E02H4.3b protein.
GN      E02H4.3b OR E02H4.3b.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC      Rhabditidae; Peloderinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Sims M.A.;
RL      Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=99069613; PubMed=9851916;
RX      none;

```

```

RT      "Genome sequence of the nematode C.elegans: A platform for
RT      investigating biology."
RL      Science 282:2012-2018(1998).
RN      [3]
RP      SEQUENCE FROM N.A.
RA      Barlow K.;
RL      Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR      EMBL; Z70205; CAD44096.1; -.
DR      EMBL; Z68003; CAD44096.1; JOINED.
DR      EMBL; Z68003; CAD44105.1; -.
DR      EMBL; Z70205; CAD44105.1; JOINED.
DR      WormPep; E02H4.3b; CE31463.
DR      InterPro; IPR000719; Prot kinase.
DR      InterPro; IPR002290; Ser_thr_kinase.
DR      InterPro; IPR001245; Tyr_kinase.
DR      Pfam; PF00069; PKINASE; 1.
DR      ProDom; PD000001; Prot kinase; 1.
DR      SMART; SM00220; S_TKc; 1.
DR      SMART; SM00219; TyTc; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.

```

DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Transferase.  
SQ SEQUENCE 409 AA; 4738 MW; F13B36A9CBA97AE CRC64;

Query Match 40.0%; Score 963; DB 5; Length 409;  
Best Local Similarity 52.2%; Pred. No. 1e-72;  
Matches 177; Conservative 64; Mismatches 94; Indels 4; Gaps 3;

QY 104 EDDEEGLICQSGDVARARVEIVDTLGEAGFAGVCEIDHGMQGMVAVKIVNGRYRE 163  
DB 54 QDDKDGHLIYSKDPFLINRTIYDTLGEFGKVR-VNDSLSDTMAIKIKNSKYRE 112  
QY 164 AARSEIQVLEHNSSTDPNSVRCVQMLEWFDHGHGVCIVFELGLSTYPIKENSPLPQ 223  
DB 113 AAKLEVYVQLQKAEKQPEKGVWVHMGSYFDYNGHICLLFDLMGSSIFDLKANHYPY 172  
QY 224 IDHIRMAYQIQSINFLHNKLTHTDLPENILFVKSIDYVVKYNSKMRDERTLKNTDI 283  
DB 173 MEQTLHTWQCNAAVFLHDKNKLTHDLKPENILFVDSRYTTLVD-KKPLRVLHSHV 230  
QY 284 KVVDPGSATYDDEHSTLVSTRHYRAPEVILALGMSQPCDWSIGCILLEYLGFVFT 343  
DB 231 RLIDFGSATFDHHSIIVSTRHYRAPEVILGMSQPCDWSIGCILLEYLGTGTLFOT 290  
QY 344 HDSKEHLLAMERILGPIPOMIQTRKRYKFFHNOIDNDHSSAGRYVRRCKPKKPYL 403  
DB 291 HENREHLLAMERVLGDIPLMARRT-KYKFINGRLDWNTSADAAVYRDNCKPLRRSMS 349  
QY 404 CHDEHEKLFEDLVRLMEYDPTQITLDEALQHPFDLL 442  
DB 350 CTDPHEVELFELIENMLMEPLARMKLPALQHRYPNRL 388

RESULT 15

035721 PRELIMINARY; PRT; 219 AA.  
ID 035721  
AC 035721  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Cdc2/CDC28-like kinase 4 (Fragment).  
GN CLK4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Watkins-Chow D.E., Douglas K.R., Buckwalter M.S., Probst F.J.,  
RA Camper S.A.;  
RT "Construction of a 3 Mb contig and transcript map of the central  
RT region of mouse chromosome 11".  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U94846; AAB62179.1; .  
DR InterPro; IPR000719; Prot. kinase.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot. kinase; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Kinase; Transferase.  
FT NON\_TER 219  
SQ SEQUENCE 219 AA; 25921 MW; CEAB6BBD2C2438C CRC64;

Query Match 34.8%; Score 839.5; DB 11; Length 219;  
Best Local Similarity 85.6%; Pred. No. 1.1e-62;  
Matches 160; Conservative 8; Mismatches 8; Indels 11; Gaps 2;

QY 8 SH-SVEEDTH-----PSHYLEASLNERDYDRRYVDYRNDYCEGYVPRHYRD 56  
DB 33 SHSSTQENRCKPHQFQSDCHYLQARCLNERDYDRRYIDYRNDYCEGYVPRHYRD 92  
QY 57 IESGYRIHCKSVSRSSPKRKRNRHCSHQSRKSHRRKRSSTIEDDEGHLICQSG 116  
:|||||

DB 93 VESTYRIHCKSVSRSSPKRKRNRPCASHQSHSGSHRRKRSSTIENDEGHLICQSG 152  
QY 117 DVLRARVEIVDTLGEAGFAGVCEIDHGMQGMVAVKIVNGRYREARARSEIQVLEHNL 176  
DB 153 DVLRARVEIVDTLGEAGFAGVCEIDHGMQGMVAVKIVNGRYREARARSEIQVLEHNL 212  
QY 177 STDPSNV 183  
DB 213 STDPSNV 219

Search completed: November 14, 2003, 15:18:41  
Job time : 41 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 03:23:10 ; Search time 7683.47 Seconds  
(Without alignments)  
19501.233 Million cell updates/sec

Title: US-10-028-946-1  
Perfect score: 6165  
Sequence: 1 atgttgagatcaaatatg.....gggaccagcttcacgataaa 6165

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estha:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hcc:\*  
9: gb\_estcl:\*  
10: gb\_estc2:\*  
11: gb\_hcc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pla:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gssi:\*  
29: gb\_gss2:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	913.6	14.8	1011	13	BQ070955 AGENCOURT
2	822.6	13.3	879	13	BQ228524 AGENCOURT
3	814	13.2	956	13	BQ071141 AGENCOURT
4	810.2	13.1	830	13	BUI81633 AGENCOURT

5	750.6	12.2	849	12	B1253509	B1253509 602973370
6	732.8	11.9	920	13	BX342268	BX342268 BX342268
7	705.6	11.4	958	11	BC031156	BC031156 Mus muscu
8	698.2	11.3	843	12	BG912161	BG912161 602812833
9	629.2	10.2	652	10	BF905370	BF905370 IL3-MT026
10	628.8	10.2	640	13	BQ087302	BQ087302 NISC_K01
11	603.4	9.8	881	12	BG976452	BG976452 602846259
12	599	9.7	891	12	B1558919	B1558919 603240969
13	591.4	9.6	746	14	CA749290	CA749290 UT-M-FY0-
14	588.6	9.5	681	9	AM605350	AM605350 QV2-DT004
15	581.6	9.4	968	13	BUI138506	BUI138506 603132657
16	575.2	9.3	1078	13	BUI139267	BUI139267 603132236
17	574.4	9.3	698	14	CD348416	CD348416 UT-M-FY0-
18	566.4	9.2	1085	12	BM904785	BM904785 AGENCOURT
19	557.4	9.0	647	10	BB207065	BB207065 BB207065
20	538	8.7	566	2	BSM071348	BSM071348 Homo sapi
21	511.8	8.3	775	13	BUI105833	BUI105833 603005490
22	511	8.3	804	13	BUI236617	BUI236617 603411670
23	498.2	8.1	719	13	BUI39218	BUI39218 603515216
24	495.6	8.0	933	13	BUI232508	BUI232508 603408272
25	488	7.9	657	13	BQ0831488	BQ0831488 L61n2149
26	481	7.8	820	13	BQ941506	BQ941506 AGENCOURT
27	451.4	7.3	459	13	BUI429033	BUI429033 UT-HF-BNO
28	450	7.3	600	12	BG808196	BG808196 2082-86 M
29	437.6	7.1	589	10	BF819093	BF819093 QV2-CI010
30	435.2	7.1	1075	13	BQ425970	BQ425970 AGENCOURT
31	431.2	7.0	497	10	BE909486	BE909486 601502838
32	424.2	6.9	662	13	BUI104008	BUI104008 603004106
33	423.8	6.9	427	12	BG944539	BG944539 ax51f11.x
34	417.2	6.8	749	13	BUI64866	BUI64866 603567832
35	416.6	6.8	678	13	BUI403714	BUI403714 604137885
36	414.8	6.7	875	13	BUI900029	BUI900029 AGENCOURT
37	412	6.7	617	13	BQ417810	BQ417810 LK51D06.Y
38	399.8	6.5	1013	13	BUI108983	BUI108983 603110947
39	396	6.4	703	14	CB512917	CB512917 ssal19b53
40	394.8	6.4	936	12	B1758420	B1758420 603026948
41	387.8	6.3	730	14	CA380902	CA380902 660300 NC
42	380.4	6.2	463	14	CB753389	CB753389 AMGNNUC:N
43	376	6.1	637	14	BY713457	BY713457 BY713457
44	374.4	6.1	522	14	T87377	T87377 Yd83e05.r1
45	373.4	6.1	473	10	BB728690	BB728690 BB728690

## ALIGNMENTS

RESULT 1  
LOCUS BQ070955  
DEFINITION AGENCOURT 6855647 NIH\_MGC\_47 Homo sapiens CDNA clone IMAGE:5923441  
5' mRNA sequence.

ACCESSION BQ070955 1011 bp mRNA linear EST 02-APR-2002  
VERSION BQ070955.1 GI:19900001  
KEYWORDS  
SOURCE  
ORGANISM

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: L1CM2091 row: p column: 02  
High quality sequence step: 634.



FEATURES  
source

Location/Qualifiers  
1. 1011  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5923441"  
/tissue\_type="neuroblastoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_47"  
/note="Organ: Brain; Vector: pORF7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library." 4 others

BASE COUNT 296 a 248 c 307 g 156 t

ORIGIN

Query Match 14.8%; Score 913.6; DB 13; Length 1011;  
Best Local Similarity 97.8%; Pred. No. 2e-206; Indels 3; Gaps 2;  
Matches 945; Conservative 0; Mismatches 18;

2701 CCGGAGGTCAGTCTAGAGCAGAGCAGAACTGAGCTCAAGCGCTCAAGAG 2760  
1 CCGGAGGTCAGTCTAGAGCAGAGCAGAACTGAGCTCAAGCGCTCAAGAG 60

2761 CTACAGCTCTCCCTGAGAGCGCGGAGTCAAGCTTGAAGCGCTGACGGGG 2820  
61 CTACAGCTCTCCCTGAGAGCGCGGAGTCAAGCTTGAAGCGCTGACGGGG 120

2821 GCCCTGAGAGCGAGCTTCCGCGAGGAGACAGAGCTGAAAGAGCAGCAGAGAG 2880  
121 GCCCTGAGAGCGAGCTTCCGCGAGGAGACAGAGCTGAAAGAGCAGCAGAGAG 180

2881 GAAGAGAGATCCAGGACTCAGGCACTAGAGATGAATCCAGGCAATTTGATGCT 2940  
181 GAAGAGAGATCCAGGACTCAGGCACTAGAGATGAATCCAGGCAATTTGATGCT 240

2941 CTTCGTACAGCTGTACTGTATACAGACCTGAGAGAGAGCTAAACAGCTGACCGAG 3000  
241 CTTCGTACAGCTGTACTGTATACAGACCTGAGAGAGAGCTAAACAGCTGACCGAG 300

3001 GACACGCTGAATCAACAACTTCTACTTGTCCAACTGATGAGGCTTCT 3060  
301 GACACGCTGAATCAACAACTTCTACTTGTCCAACTGATGAGGCTTCT 360

3061 GGGCGCAAGCAGAGTTGACAACTGCGAGAGTGAAGTGAACATCTCCGCGGAGAGATC 3120  
361 GGGCGCAAGCAGAGTTGACAACTGCGAGAGTGAAGTGAACATCTCCGCGGAGAGATC 420

3121 ACCGAAAGAGATGAGCTTACACAGCAGAAACAGATGAGAGCTTGAAGACAG 3180  
421 ACCGAAAGAGATGAGCTTACACAGCAGAAACAGATGAGAGCTTGAAGACAG 480

3181 TGACACATGCTGAGAGAAAGCTGATGAGTTTGAAGCCCTTAAACATGAGCTGTAGA 3240  
481 TGACACATGCTGAGAGAAAGCTGATGAGTTTGAAGCCCTTAAACATGAGCTGTAGA 540

3241 AAAGAGCGGAGTGGAGGCTTGAAGAGAGCTCTGAGTGAAGAAATCCAGCTTTGAG 3300  
541 AAAGAGCGGAGTGGAGGCTTGAAGAGAGCTCTGAGTGAAGAAATCCAGCTTTGAG 600

3301 TGTGCGGTTTGAAGAGCTGCAAGAGATGCTGAGACCGAAGAACAGAGCGGCGAGAGCC 3360  
601 TGTGCGGTTTGAAGAGCTGCAAGAGATGCTGAGACCGAAGAACAGAGCGGCGAGAGCC 660

3361 GATCAGCGAGTCAAGAGCTCTGCGCAGAGTGTGAGAGCTGAGTGAAGAGCAGAGGCT 3420  
661 GATCAGCGAGTCAAGAGCTCTGCGCAGAGTGTGAGAGCTGAGTGAAGAGCAGAGGCT 720

3421 GAGATTCTCGCTCGAGAGGCTCTCAAGAGCAGAGTGAAGGCCGAGAGCTCTCT 3480  
721 GAGATTCTCGCTCGAGAGGCTCTCAAGAGCAGAGTGAAGGCCGAGAGCTCTCT 780

3481 GACAGGCTCAATGACCTGAGAGAGAGCAGATGATGCTTGAATGAATGCC- GAAGCTT 3539  
781 GACAGGCTCAATGACCTGAGAGAGAGCAGATGATGCTTGAATGAATGCCGAGAGCTT 840

3540 ACAGAGAGCTGAGAGCTGAGAGAGCTGAGAGAGCTTGAAGAGAGCAGAGCAG 3599  
841 ACAGAGAGCTGAGAGCTGAGAGAGCTGAGAGAGCTTGAAGAGAGCAGAGCAG 900

3600 ATTACAGAGAGATGAGAGCTGAGAGAGCTTGAAGAGAGCTTGAAGAGAGCTG 3657  
901 ATTACAGAGAGATGAGAGCTGAGAGAGCTTGAAGAGAGCTTGAAGAGAGCTG 960

3658 CAGAGA 3663  
961 CAGAGA 966

RESULT 2  
BQ228524  
LOCUS  
DEFINITION  
AGENCOURT 7600932 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:5061485  
5', mRNA sequence.  
BQ228524  
BQ228524.1 GI:20409924  
EST.  
SOURCE  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 879)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DP  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: L1AM1331 row: 9 column: 22  
High quality sequence stop: 636.

FEATURES  
source  
Location/Qualifiers  
1. 879  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5061485"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_72"  
/note="Organ: skin; Vector: pCMV-Sport6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies." 1 others

BASE COUNT 221 a 258 c 239 g 160 t 1 others

ORIGIN

Query Match 13.3%; Score 822.6; DB 13; Length 879;  
Best Local Similarity 98.0%; Pred. No. 9.1e-185;  
Matches 853; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

3949 CCGGAGAGAGCTGCGCCACCGCAAGCAGCAGACCAACCAACCAACCAACCAAC 4008  
1 CCGGAGAGAGCTGCGCCACCGCAAGCAGCAGACCAACCAACCAACCAACCAAC 60

4009 GCGAGCAGCAGATTCGCAATGTCGCAATGTCGCGATGCGCAGAGCAGCAGGCTGC 4068



Db	541	AAAGAGCGGACGATGGAGAGGCTCTGGAGAGCGCTCTGGGTGATGAGAAATCCAGTTTGAG	600
OY	3301	TGTCGGGTTTCAGAGGCTGCAGAGGATGCTGTGACACCCGAGAAACAGAGAGGCGCAGAGCC	3360
Db	601	TGTGGGTTTCGAGACTCCAGAGAAATGCTGTGACACCGAGAAACAGAGAGGCGCAGAGCC	660
OY	3361	GATCAGCGGATCACCGAGTCTTCGCCAGGTGTGTGAGCTTGCGCAGTGAAGAGACACAGAGCT	3420
Db	661	GATCAGCGGATCACCGAGTCTTCGCCAGGTGTGTGAGCTTGCGCAGTGAAGAGACACAGAGCT	720
OY	3421	GAGATTTCTCGCTCTGACAGGCTCTCAAAGAGCGAGAAAGCTGAAGGCCGAGAGCCTCTCT	3480
Db	721	GAGATTTCTCGCTCTGACAGGCTCTCAAAGAGCGAGAAAGCTGAAGGCCGAGAGCCTCTCT	780
OY	3481	GACCAAGCTCAATGACTT-GGAGAAAGACATGCTATGCTTGAAA--TGAATGCCCGAAGC	3537
Db	781	GACCAAGCTCAATGACTTGGGAGAAAGAACATGCTATGCTTGAAAATGAATGCCCGAAGC	840
OY	3538	TTACAGCGAGAA--GCTGGAGACTGAAC--GAGAGCTCAAAACAGAGGCTTTCTGGAAAGACCA	3592
Db	841	TTACAGCGAGAAAGCTGGGAGACTGAACCCGAGAAAGCTTCAAAACAGAGGCTTTCTGGAGAGGA	900

RESULT 4	830 bp	mRNA	linear	EST 04-SEP-2002
BUI181633				
LOCUS				
DEFINITION				
BUI181633				
AGENCOURT 7906221 NIH_MGC_67 Homo sapiens CDNA clone IMAGE:6140538				
5', mRNA sequence.				
ATTGATGCTG				

VERSION	BU181633.1	GI:22695617
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (basses 1 to 830)  
TITLE NIH-MGC <http://mgc.nci.nih.gov/>.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished  
Contact: Robert Strausberg, Ph.D.

115356 Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LRM3459 row: e column: 19  
 High quality sequence, stop: 652.

FEATURES	
SOURCE	LOCATION/VARIABLES
1. .830	
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="IMAGE:6140538"
	/tissue_type="retinoblastoma"
	/lab_host="DH10B (phage-resistant)"
	/clone_1lb="N1H.MGC.67"
	/note="Organ: eye; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."
BASE COUNT	259 a 202 c 239 g 129 t 1 others
ORIGIN	

Query Match	13.1%	Score 810.2	DB 13	Length 830
Best Local Similarity	99.4%	Pred. No. 8e-182		
Matches 823	Conservative 0	Mismatches 4	Indels 1	Gaps 1
OY	2435	TGATTCAGATCAGATCCCTGGAAACAGAGGATTTGGAACCTGCTCGAACCCATAAAC		2434
Db	1	TGGATTCGAAATAGATCCCTGGAAACAGAGGATTTGGAACCTGCTCGAACCCATAAAC		60

OY	2495	TTGCAGCAAAATAGAGCTCTTTTACCCAAAGAACTAGAGGCCCAAGACAGATGATT	2554
Db	61	TTGCAGCAAAATAGAGCTCTTTTACCCAAAGAACTAGAGGCCCAAGACAGATGATT	120
OY	2555	CTGAACCTAGGCAACAGAAATTTTAACTGGAGACACAGCTGGGAAAGTTGGAGGCCAGA	2614
Db	121	CTGAACCTAGGCAACAGAAATTTTAACTGGAGACACAGGCTGGGAAAGTTGGAGGCCAGA	180
OY	2615	ACCGAAAACCTGGAGAGAGAGCTGGAGAAAGATCAGCGAACCAAGACACAGCTGACAAAGATC	2674
Db	181	ACCGAAAACCTGGAGAGAGAGCTGGAGAAAGATCAGCGAACCAAGACACAGCTGACAAAGATC	240
OY	2675	GAGCTGCTGAACTGGAGACAAAGATTGCGGAGGTGAGTCTTAGAGTACGAGAGCGAAGAAC	2734
Db	241	GAGCTGCTGAACTGGAGACAAAGATTGCGGAGGTGAGTCTTAGAGTACGAGAGCGAAGAAC	300
OY	2735	TGGAGCTCAAGCGGCGAGCTCAACAGAGCTACAGCTCTCCCTGACGAGGCGGAGTACAGT	2794
Db	301	TGGAGCTCAAGCGGCGGAGCTCAACAGAGCTACAGCTCTCCCTGACGAGGCGGAGTACAGT	360
OY	2795	TGACAGCCCTGACAGCTGACAGCGGCGGCGCTTGAGAGAGCCAGCTTTCGACGGCGAAGACAG	2854
Db	361	TGACAGCCCTGACAGCTGACAGCGGCGGCGCTTGAGAGAGCCAGCTTTCGACGGCGAAGACAG	420
OY	2855	AGCTGGAAGAGACACACAGCAGCAAGCTGGAAGAGAGATCAGGACCTCACGGGACATAGAG	2914
Db	421	AGCTGGAAGAGACACACAGCAGCAAGCTGGAAGAGAGATCAGGACCTCACGGGACATAGAG	480
OY	2915	ATGAAATCAGCGCAAAATTGATGCTCTTGTAACAGCTGATCTGTAAATCAACAGCCTGG	2974
Db	481	ATGAAATCAGCGCAAAATTGATGCTCTCTGTAACAGCTGATCTGTAAATCAACAGCCTGG	540
OY	2975	AGGAGCAGCTAAACACAGCTGACCGAGGACAAAGCTGAACTCAACACAAAACCTTCTACT	3034
Db	541	AGGAGCAGCTAAACACAGCTGACCGAGGACAAAGCTGAACTCAACACAAAACCTTCTACT	600
OY	3035	TGTCGAAAACCTCGATGAGGCTTTCTGGCGGCAACGACGAGATTGTCAACTGCCAAGTGG	3094
Db	601	TGTCGAAAACCTCGATGAGGCTTTCTGGCGGCAACGACGAGATTGTCAACTGCCAAGTGG	660
OY	3095	AAGTGGACCAATCTCCGCGGGAGATTCACGAGACGAGAGATGACAGTTTACCGACGCGAAGGC	3154
Db	661	AAGTGGACCAATCTCCGCGGGAGATTCACGAGACGAGAGATGACAGTTTACCGACGCGAAGGC	720
OY	3155	AAACGATGAGGCTCTGAGACCAAGTGCACCAATCTTGGAGGAAACAGGTCAATGAGATT - G	3213
Db	721	AAACGATGAGGCTCTGAGACCAAGTGCACCAATCTTGGAGGAAACAGGTCAATGAGATTG	780
OY	3214	GAGGCGCTTAACGATGAGCTGCTGAAAAAAGGCGGACAGTGGGAGGCC 3261	
Db	781	GAGGCGCTTAACGATGAGCTGCTGAAAAAAGGCGGACAGTGGGAGGCC 828	

RESULT	5
LOCUS	B1253509
DEFINITION	B1253509 849 bp. mRNA linear EST 17-JUL-2001
ACCESSION	602973730P1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5112737 5', mRNA sequence.
VERSION	B1253509
KEYWORDS	B1253509..1 GI:14805003 EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 849) NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished
AUTHORS	Contact: Robert Strausberg, Ph.D. Email: cgapds-team1.nih.gov
TITLE	
JOURNAL	
COMMENT	Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: InCyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
 http://image.llnl.gov  
 Plate: L1AM11273 row: 1 column: 18  
 High quality sequence start: 2  
 High quality sequence stop: 772.

## FEATURES

## SOURCE

1.849  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5112737"  
 /tissue\_type="cervical carcinoma cell line"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_12"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.4 kb. Library prepared by Life  
 Technologies."

BASE COUNT 218 a 230 c 211 g 190 t  
 ORIGIN

Query Match 12.2%; Score 750.6; DB 12; Length 849;  
 Best Local Similarity 98.5%; Pred. No. 1.3e-167;  
 Matches 789; Conservative 0; Mismatches 9; Indels 3; Gaps 3;

QY 4812 ACTGCTTGAAGACCTCCCTGCTGAAGCTGGAAGGATGATACCGCTTGAACATGAACTGCAC 4871  
 DB 3 ACGCGTGAAGACCTCCCTGCTGAAGCTGGAAGGATGATACCGCTTGAACATGAACTGCAC 62  
 QY 4872 GCTGCTTGAAGACCTCCCTGCTGAAGCTGGAAGGATGATACCGCTTGAACATGAACTGCAC 4931  
 DB 63 GCTGCTTGAAGACCTCCCTGCTGAAGCTGGAAGGATGATACCGCTTGAACATGAACTGCAC 122  
 QY 4932 TGTCTTGAAGACCTCCCTGCTGAAGCTGGAAGGATGATACCGCTTGAACATGAACTGCAC 4991  
 DB 123 TGTCTTGAAGACCTCCCTGCTGAAGCTGGAAGGATGATACCGCTTGAACATGAACTGCAC 182  
 QY 4992 TATCAAGACCTGGAAGACCTGATGATGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5051  
 DB 183 TATCAAGACCTGGAAGACCTGATGATGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242  
 QY 5052 GGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5111  
 DB 243 GGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302  
 QY 5112 CTCACCAACATTTTGAAGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5171  
 DB 303 CTCACCAACATTTTGAAGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362  
 QY 5172 GAACGAGAGCTGTCAGATCTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5231  
 DB 363 GAACGAGAGCTGTCAGATCTGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 422  
 QY 5232 AAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5291  
 DB 423 AAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482  
 QY 5292 CCACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5351  
 DB 483 CCACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 542  
 QY 5352 GCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5411  
 DB 543 GCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 602  
 QY 5412 TGCGGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5471  
 DB 603 TGCGGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662

QY 5472 AGAGAGTACTTGTCTGTGTTTCCAGATTTGAGAGTGTGTGATTTCTTACGAGAGAG 5531  
 DB 663 AGAGAGTACTTGTCTGTGTTTCCAGATTTGAGAGTGTGTGATTTCTTACGAGAGAG 722  
 QY 5532 T-AGCGGACAGAGAGATCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 5590  
 DB 723 T-AGCGGACAGAGAGATCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 780  
 QY 5591 CCTATCTGTTTGTGACCCACT 5611  
 DB 781 CCTATCTGTTTGTGACCCACT 801

RESULT 6  
 BX342268  
 LOCUS  
 DEFINITION BX342268 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens  
 ACCESSION BX342268  
 VERSION BX342268  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 920)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 6533.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DK011YB201&cluster=6533.r. Contact :  
 Feng Liang Email: fliang@life.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0DK011YB201.

## FEATURES

## SOURCE

1.920  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DK011YB20"  
 /cell\_type="HELA CELLS COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and BcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 BASE COUNT 216 a 298 c 249 g 157 t  
 ORIGIN

Query Match 11.9%; Score 732.8; DB 13; Length 920;  
 Best Local Similarity 99.6%; Pred. No. 2.4e-163;  
 Matches 745; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 5418 CTCCTCCAAAGAGCTCCCTGCTCAATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5477  
 DB 66 CTCCTCCAAAGAGCTCCCTGCTCAATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 125  
 QY 5478 GTACTGTGTGTTTCCAGAGATTTGAGAGTGTGTGATTTCTTACGAGAGAGAGAGAGAGAG 5537  
 DB 126 GTACTGTGTGTTTCCAGAGATTTGAGAGTGTGTGATTTCTTACGAGAGAGAGAGAGAGAG 185  
 QY 5538 CACGAGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5597  
 DB 186 CACGAGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245





QY	6016	GAGAAAGTCCCCCGGCGCGATCTCAGCACACGCGAGAGAGCGGTCCCCCGCAGAGGCTGTTT	6075
Db	760	GHGANGTCCCCAGGCGCGAAATGCTCAGCACACTTGGAGAGAGCGGTCCCCAGAGACTGTTT	819
QY	6076	GAAGCACACAGCAGGGGCGCGGTGCTTCGCGAGAGCGCTGAGAGACCCCGCTGTCCACAGGTG	6135
Db	820	GAAGCACACAGCAGGGGCGCGGTGCTTCAGAGAGCAGTGAAGACCCCACTGTCCAGGTT	879
QY	6136	AACAAAGTCTGGAGACCACTTCACTATTA	6165
Db	880	AACAAAGTCTGGAGACCACTTCACTATTA	909

RESULT 8  
LOCUS BG912161  
DEFINITION BG912161 843 bp mRNA linear EST 05-JUN-2001  
6028128331 NCI\_CGAP\_Brn67 Homo sapiens cDNA clone IMAGE:4944657  
5', mRNA sequence.  
ACCESSION BG912161  
VERSION BG912161.1 GI:14292637  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 843)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, ph.D.

Tissue Procurement: David N. Louis, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at:  
<http://image.llnl.gov>  
 Plate: LIML10890 row: a column: 10  
 High quality sequence stop: 778.

FEATURES	
source	Location/Qualifiers
1..843	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="taxon:9606"	
/clone IMAGE:4944657"	
/issue_type="anaplastic oligodendroglioma with 1p/19q loss"	
/lab_host="DH10B (T1 phage-resistant)"	
/clone_1b="NCI CGAP Brn67"	
/note="Organ: brain; Vector: pCMV-Sport6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."	
218 a	
227 c	
214 g	
164 t	

	Query Match	Best Local Similarity	Matches	785; Conservative	0; Mismatches	48; Indels	8; Gaps	5; DB 12; Length 843; Score 698.2; Pred. No. 4.2e155;
QY	4669	CACCCGCACACCACTCTGTGGCCCGGAGAAACCTTACTTGTCTAGTCTCCAGCTTCCCT	4728					
Db	2	CACCGTCGACCACTCTGTGGCCCGGAGAAACCTTACTTGTCTAGTCTCCAGCTTCCCT	61					
QY	4729	GACAAAGAGCGCTGGGTCAACCGCCCTTGAATACATTTGTGCGAGGTGGAGAGTTTCTAGG	4788					
Db	62	GACAAAGAGCGCTGGGTCAACCGCCCTTGAATACATTTGTGCGAGGTGGAGAGTTTCTAGG	121					
QY	4789	GAAAGACGAGAGCTGATGTCTAAACTGCTTGAAGTCCCTGCGGAACTGGAAGGTGAT	4848					
Db	122	GAAAGACGAGAGCTGATGTCTAAACTGCTTGAAGTCCCTGCGGAACTGGAAGGTGAT	181					

QY	4849	GACCGTTTACACATGAACTGACAGCTGACCTTCAAGTACACAGATGGTGTGGGACC	4908
Dp	182	GACCGTTTACACATGAACTGACAGCTGACCTTCAAGTACACAGATGGTGTGGGACC	241
QY	4909	GAGCAAGAGCTCTACGCCCTGAATGTCTTGGAAAAATCTCCCTAACCCATGTCCAGGAATT	4968
Dp	242	GAGCAAGAGGCTCTACGCCCTGAATGTCTTGGAAAAATCTCCCTAACCCATGTCCAGGAATT	301
QY	4969	GGAGCAGTCTTCCAAATTTATATATATCAAGAGACCTGAGAGAGTACTCATGATGAGAGGA	5028
Dp	302	GGAGCAGTCTTCCAAATTTATATATATCAAGAGACCTGAGAGAGTACTCATGATGAGAGGA	361
QY	5029	GAAAGAGCGGACACTGTGTCTTGTGACGTGAAAGAAAGTAAACAGTCCCTGGCCAGTCC	5088
Dp	362	GAAAGAGCGGACACTGTGTGTGTGGACGTGAAAGAAAGTAAACAGTCCCTGGCCAGTCC	421
QY	5089	CACCTGCTCTGCCAGGCCGACATCTGACCCCAATTTTGAAGCTGTCAAGGGCTGCAC	5148
Dp	422	CACCTGCTCTGCCAGGCCGACATCTGACCCCAATTTTGAAGCTGTCAAGGGCTGCAC	481
QY	5149	TGTGTTGGGGCAGGCAAGATTGAGAACGGGCTCTGCAATCTGTGACAGCCATGCCAGCAAA	5208
Dp	482	TGTGTTGGGGCAGGCAAGATTGAGAACGGGCTCTGCAATCTGTGACAGCCATGCCAGCAAA	541
QY	5209	GTCGTCAATTCCTCGGCTACACGAAACCTGACAAATPACGTACCTCGGAAAGATAGAG	5268
Dp	542	GTCGTCAATTCCTCGGCTACACGAAACCTGACAAATPACGTACCTCGGAAAGATAGAG	601
QY	5269	ACCTCAGAGGCCCTGACAGCTGTATCCACTTCAACCAATTACAGTATCTCTAAT--GGAAACAA	5327
Dp	602	ACCTCAGAGGCCCTGACAGCTGTATCCACTTCAACCAATTACAGTATCTCTAATGGAAACAA	661
QY	5328	TAAATTCCTACG-AAATTCGACATGAA-GCAGTACAGCTCGAGGAAT--CCTGAGATAGG	5383
Dp	662	TAAATTCCTACGAAATTCGACATGAAAGCAGTACACGCTCGAGGAATTTCTGTGATACGA	721
QY	5384	ATGACCAATTCCTTGGCAGCTGCTGTG--TTTGCGGCTCTTCCACAGCTTCCCTGTCT	5440
Dp	722	ATGACCAATTCCTTGGGACCTGAGCTGGATTACCCCTCTTCCAAAAGTTCTGTCTCTC	781
QY	5441	CAATTCGGACGATGAACAGCGACGGGGAAGGAGAGATGACTTGTGTTTCCAGGAAT	5500
Dp	782	AAATTCGGACGATGAACAGCGACGGGGAAGGAGATGATTGCTGTGTTCCTCCCGACT	841
QY	5501	T 5501	
Dp	842	T 842	

[illegible]

Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/ICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&t2=IL3-MT0267-  
261200-410-H07&t3=2000-12-26&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 624.  
Location/Qualifiers  
1. .652  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="MT0267"  
/note="Organ: marrow; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 94 a 189 c 163 g 206 t  
ORIGIN

Query Match 10.2%; Score 629.2; DB 10; Length 652;  
Best Local Similarity 98.0%; Pred. No. 1e-138;  
Matches 637; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 2525 GGAACATGAAGGCCCAAGAGAGATGATTTCTGAAGTCAAGGCAAGAAATTTTACCTGG 2584  
DB 652 GGAACATGAAGGCCCAAGAGAGATGATTTCTGAAGTCAAGGCAAGAAATTTTACCTGG 593  
QY 2585 AGACACAGGCTGGGAAAGTTGGAGGCCCAAGAACCAAACTGAGAGACGAGCTGAGAGA 2644  
DB 592 AGACACAGGCTGGGAAAGTTGGAGGCCCAAGAACCAAACTGAGAGACGAGCTGAGAGA 533  
QY 2645 TCAGCCCAAGAGCAAGTGAACAAGATGGGCTGCTGGAATCTGGAGACAAAGATTTGGGG 2704  
DB 532 TCAGCCCAAGAGCAAGTGAACAAGATGGGCTGCTGGAATCTGGAGACAAAGATTTGGGG 473  
QY 2705 AGGTCACTTAGAGCAAGAGAGAGAGAACTGAGCTCAAGCGCCAGCTCAAGAGCTAC 2764  
DB 472 AGGTCACTTAGAGCAAGAGAGAGAGAACTGAGCTCAAGCGCCAGCTCAAGAGCTAC 413  
QY 2765 AGCTCTCCCTGCAAGAGCGGAGTCAAGTTGACAGCCCTGCAAGGCTGCAACGGGCGGCC 2824  
DB 412 AGCTCTCCCTGCAAGAGCGGAGTCAAGTTGACAGCCCTGCAAGGCTGCAACGGGCGGCC 353  
QY 2825 TGGAGAGCCAGCTTCGCGCAAGGCGAAGACAGAGCTGGAAGAGACCAAGCAGAGAGCTGAAG 2884  
DB 352 TGGAGAGCCAGCTTCGCGCAAGGCGAAGACAGAGCTGGAAGAGACCAAGCAGAGAGCTGAAG 293  
QY 2885 AGAGATCCAGGCACTTCACGCGCAATAGAGATGAATTCAGCGCAAAATTTGATGCTCTTC 2944  
DB 292 AGAGATCCAGGCACTTCACGCGCAATAGAGATGAATTCAGCGCAAAATTTGATGCTCTTC 233  
QY 2945 GTAAAGCTGTACTGTAAATCAAGACCTGAGAGAGCGCTAAACCAAGCTGACCGAGACA 3004  
DB 232 GTAAAGCTGTACTGTAAATCAAGACCTGAGAGAGCGCTAAACCAAGCTGACCGAGACA 173  
QY 3005 ACAGCTGAATCAACAACAATCTTCTACTGTCCAAACAACCTGATGAGGCTTTGGGG 3064  
DB 172 ACAGCTGAATCAACAACAATCTTCTACTGTCCAAACAACCTGATGAGGCTTTGGGG 113  
QY 3065 CCAACGACGAGATTGTACAACTCGGAGAGTGAAGTGAACATCTTCGCGCGGAGATCAACG 3124  
DB 112 CCAACGACGAGATTGTACAACTCGGAGAGTGAAGTGAACATCTTCGCGCGGAGATCAACG 53

QY 3125 AACGAGAGATGACAGCTTACGACCCAGAGACAAACGATGAGGCTGTGAG 3174  
DB 52 AACGAGAGATGACAGCTTACGACCCAGAGACAAACGATGAGGCTGTGAG 3  
RESULT 10  
BO807302 640 bp mRNA linear EST 31-JUL-2002  
LOCUS NISC\_KK01b10.y1 NCI CGAP\_Brn72 Macaca mulatta cDNA clone  
DEFINITION IMAGE5330250.5', mRNA sequence.  
ACCESSION BO807302  
VERSION BO807302.1 GI:22031511  
KEYWORDS EST.  
SOURCE Macaca mulatta (rhesus monkey)  
ORGANISM Macaca mulatta  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
Cercopitheciinae; Macaca.  
REFERENCE 1 (bases 1 to 640)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
CDNA Library Preparation:  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/ILNL  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC)  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNL at:  
info@image.llnl.gov  
Plate: LLAM1838 row: C column: 19  
Seq primer: M13RPI reverse primer (ABI).  
Location/Qualifiers  
1. .640  
/organism="Macaca mulatta"  
/mol\_type="mRNA"  
/db\_xref="taxon:9544"  
/clone\_lib="IMAGE:5330250"  
/tissue\_type="hypothalamus"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NCI CGAP Brn72"  
/note="Organ: brain; Vector: PCMV-SPOrte.ccd; Site\_1:  
NotI; Site\_2: EcoRV; Cloned unidirectionally. Primer:  
Oligo dT. Average insert size 2.2 kb. Constructed by  
Invitrogen. Note: this is a NCI-CGAP Library."

BASE COUNT 166 a 182 c 166 g 126 t  
ORIGIN

Query Match 10.2%; Score 628.8; DB 13; Length 640;  
Best Local Similarity 98.9%; Pred. No. 1.3e-138;  
Matches 633; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 4073 GCGTGTGGCCCGCCATCCAGCCGAGAAAGAGTCTTAAGTCCAGAGAAATTTAGTC 4132  
DB 1 GCGTGTGGCCCGCCATCCAGCCGAGAAAGAGTCTTAAGTCCAGAGAAATTTAGTC 60  
QY 4133 GCGGTCTTAAAGAACCATGACCAATATTTCTCAACGATTCACAGTGAAGTGAACA 4192  
DB 61 GCGGTCTTAAAGAACCATGACCAATATTTCTCAACGATTCACAGTGAAGTGAACA 120  
QY 4193 TGGAGCCCAAGATGTGCTGTGTCTGGATACCGGTGACCTTGGAGCGCCAGGATCCA 4252  
DB 121 TGGAGCCCAAGATGTGCTGTGTCTGGATACCGGTGACCTTGGAGCGCCAGGATCCA 180  
QY 4253 AATGTCTGAATGACAGTGAATGTGTACCCCAAGTCTCAAGTGTTCACCAAGCACT 4312  
DB 181 AATGTCTGAATGACAGTGAATGTGTACCCCAAGTCTCAAGTGTTCACCAAGCACT 240  
QY 4313 GCGGCTTGGCTGTGAATATGACACACTTTCACCAAGGCTTTCGCGGAGAAATGA 4372



Db 241 GCGGCTGCTGCGGATACGACACACTTCTAGAGGCTTCTGCGGACAAATGA 300  
QY 4373 ACTCCAGAGTCTCCAGACCAAGAGCCAGCAGCAGCTTGCACCTGAAAGGATGA 4432  
Db 301 ACTCCAGAGTCTCCAGACCAAGAGCCAGCAGCAGCTTGCACCTGAAAGGATGA 360  
QY 4433 AGGTGCGGAGATACAAAGAGAGAGAGAGCTGAGAGAGATGATCTGCTGG 4492  
Db 361 AGGTGCGGAGATACAAAGAGAGAGAGAGCTGAGAGAGATGATCTGCTGG 420  
QY 4493 AGGATACAAAGTCTCTATTTATGACAAATGAGAGAGAGAGAGAGAGAGAG 4552  
Db 421 AGGATACAAAGTCTCTATTTATGACAAATGAGAGAGAGAGAGAGAGAGAG 480  
QY 4553 AAGATTTGAGCTGCTGCTTCCGAGAGAGAGATGATCTATTCATGAGTCCGTTGCTT 4612  
Db 481 AAGATTTGAGCTGCTGCTTCCGAGAGAGAGATGATCTATTCATGAGTCCGTTGCTT 540  
QY 4613 CCGAAGTCCGAATATACAGCAAGAGAGAGATGATCTATTCATGAGTCCGTTGCTT 4672  
Db 541 CCGAAGTCCGAATATACAGCAAGAGAGAGATGATCTATTCATGAGTCCGTTGCTT 600  
QY 4673 CCGAAGTCCGAATATACAGCAAGAGAGAGATGATCTATTCATGAGTCCGTTGCTT 4712  
Db 601 CCGAAGTCCGAATATACAGCAAGAGAGAGATGATCTATTCATGAGTCCGTTGCTT 640

RESULT 11  
Bg976452 881 bp mRNA linear EST 12-JUN-2001  
LOCUS 602846269F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:4976752 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg976452 GI:14364089  
VERSION Bg976452.1 GI:14364089  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 881)  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: CGAPsb-rt@mail.nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM10970 row: j column: 17  
High quality sequence stop: 841.  
Location/Qualifiers  
1..881  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4976752"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Mam6"  
/note="Organ: mammary; Vector: pCMV-Sport6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 212 a 269 c 241 g 158 t 1 others  
ORIGIN

Query Match 9.8%; Score 603.4; DB 12; Length 881;  
Best Local Similarity 86.2%; Pred. No. 1.7e-132;  
Matches 726; Conservative 0; Mismatches 107; Indels 9; Gaps 5;

QY 5302 AATTACGATTCCTCATTTGAAACCAATTAATTCTAGAAATCG/CATGAGCAGTACAG 5361  
Db 40 AAGAAAGCATCTCTCATTTGGAGCAACAAATTTATGAGATGACATGAGCAGTACAG 99  
QY 5362 CTGAGGAATTCCTGATTAAGATGACATTCCTTGGACCTGCTGTTTGCCTCTG 5421  
Db 100 CTGATGATGTTCTTGACAAAGAGAGACATTCCTTGGACCTGCTGTTTGCCTCTG 159  
QY 5422 TCCAAAGCTTCTCTGCTGATTCGAGTGAACAGCGCAGGCGACGAGAGACTAC 5481  
Db 160 TCCAAAGCTTCTCTGCTGATTCGAGTGAACAGCGCAGGCGACGAGAGACTAC 219  
QY 5482 TTGCTGCTTCCAGCAATTTGAGAGTTGCTGATTTCTTACGAAAGAGTACCGGACA 5541  
Db 220 CTGCTGCTTCCAGCAATTTGAGAGTTGCTGATTTCTTACGAAAGAGTACCGGACA 279  
QY 5542 GACGATCTCAAGTGAAGTCTGATTCCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 5601  
Db 280 GATGATCTTAAGTGAAGTCTGATTCCTTGGCTTGGCTTGGCTTGGCTTGGCTTGG 339  
QY 5602 GTGACCACTTCAACTACTGCAAGTATGAGATTCAGGCAAGCTTCTTACAGAGGAC 5661  
Db 340 GTGACCTACTTCAACTCTCGAAGTATGAGATTCAGGCAAGCTTCTTACAGAGGAC 399  
QY 5662 CTTGCGCGAGCAGTACCTGACATTCGCAACCCCGCTTACCTGAGGCTTGCCTTCTCA 5721  
Db 400 CTTGCGCGAGCAGTACCTGACATTCGCAACCCCGCTTACCTGAGGCTTGCCTTCTCA 459  
QY 5722 GAGCGATTTACTTGGCTCTCATACAGAGATTAAGGTCATTGCTGCAAGGA 5781  
Db 460 GAGCGATTTACTTGGCTCTCATACAGAGATTAAGGTCATTGCTGCAAGGA 539  
QY 5782 AACCTGTGAAGAGTCCGCACTGAACACACCCGCGGCGCTTCCACCTCCGAGAGAG 5841  
Db 520 AACCTGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 579  
QY 5842 CCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5901  
Db 580 CCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638  
QY 5902 CCGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5956  
Db 639 CCGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 697  
QY 5957 ACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6016  
Db 698 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757  
QY 6017 AAGAGTCTCCCGCGCGAGTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6074  
Db 758 AAGAGTCTCCCGCGCGAGTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 817  
QY 6075 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6134  
Db 818 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 877  
QY 6135 GA 6136  
Db 878 AA 879

RESULT 12  
B1558919 891 bp mRNA linear EST 05-SEP-2001  
LOCUS B1558919  
DEFINITION 603240969F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5293913 5',  
mRNA sequence.  
ACCESSION B1558919  
VERSION B1558919.1 GI:15446233  
KEYWORDS EST.



size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCAGACAG. This library was created for the University Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 209 a 185 c 243 g 106 t 3 others  
 Query Match 9.6%; Score 591.4; DB 14; Length 746;  
 Best Local Similarity 88.5%; Pred. No. 1.1e-129;  
 Matches 662; Conservative 0; Mismatches 84; Indels 2; Gaps 2;

3085 CTGCGAAGTGAAGTGAACATCTCCGCGGAGATCAAGAGAGATGACACTTACC 3144  
 1 CTGCGAGTGAAGTGAACATCTCCGCGGAGATCAAGAGAGATGACACTTACC 60  
 3145 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3204  
 61 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
 3205 ATGGAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3264  
 121 CTGAGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
 3265 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3324  
 181 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
 3325 ATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3384  
 241 ATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
 3385 CAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3444  
 301 CAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
 3445 CTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3504  
 361 CTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
 3505 AAGCATGCTATGCTTGAATGAATGCCGAAGCTTACAGCAAGAGCTGAGACGA 3564  
 421 AAGCATGCTATGCTTGAATGAATGCCGAAGCTTACAGCAAGAGCTGAGACGA 480  
 3565 GAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3624  
 481 GAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 3625 AAAAATCAATTTCCGCTCTGACTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 3684  
 541 AAGAAATCAATTTCCGCTCTGACTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 3685 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3744  
 601 CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
 3745 CATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3804  
 661 CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719  
 3805 CAAGCAAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3832  
 720 CAAG-CANAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 746

RESULT 14  
 AM605350/c 681 bp mRNA linear EST 23-MAR-2000  
 LOCUS  
 DEFINITION QV3-DT0045-210100-063-h08 DT0045 Homo sapiens cDNA, mRNA sequence.

ACCESSION AM605350  
 VERSION GI:7310091  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 681)  
 HCGP <http://www.ludwig.org.br/ORESTES>.  
 TITLE The FAPESP/LICR Human Cancer Genome Project  
 JOURNAL Unpublished  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV3&t2=QV3-DT0045-210100-063-h08&t3=2000-01-21&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 20  
 Location/Qualifiers

FEATURES  
 Source  
 1..681  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /date="1996-06-06"  
 /clone\_lib="DT0045"  
 /note="Origin: denis drash; Vector: puc18; Site: 1; Smal; Site 2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 141 a 171 c 190 g 178 t 1 others  
 ORIGIN

Query Match 9.5%; Score 588.6; DB 9; Length 681;  
 Best Local Similarity 98.7%; Pred. No. 5.1e-129;  
 Matches 624; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

4643 TCCCATCATCTGAAGATGAGATCTCACCCGACACCACTGTGCGCCGGAGAACCC 4702  
 634 TCCCATCATCTGAAGATGAGATCTCACCCGACACCACTGTGCGCCGGAGAACCC 577  
 4703 TCTACTGCTAGCTCCAGCTTCCCTGACAAACAGCGCTGGGTACCGCTTGAATCG 4762  
 576 TCTACTGCTAGCTCCAGCTTCCCTGACAAACAGCGCTGGGTACCGCTTGAATCG 517  
 4763 TTGTGCGAGGTGGAGAGATTTCTAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 4822  
 516 TTGTGCGAGGTGGAGAGATTTCTAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 457  
 4823 ACTCCCTGCTGAAACTGGAAGGTGAGACCGTCTAGAGATGACGAGCTGCTTCA 4882  
 456 ACTCCCTGCTGAAACTGGAAGGTGAGACCGTCTAGAGATGACGAGCTGCTTCA 397  
 4883 GTGACGAGGTGTTGGTGGGACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4942  
 396 GTGACGAGGTGTTGGTGGGACACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 337  
 4943 ACTCCCTAACCATGCTCCAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 5002  
 336 ACTCCCTAACCATGCTCCAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 277  
 5003 TGGAGAGCTACTCATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5062

Db 276 TGGAGAGCTACTCATGATAGCAGAGAGAGCGGCACTGTGCTTGTGGACGTGAAGA 217

QY 5063 AATGTAAGCTGCTCCGCGCCAGTCCCACTGCTGCCAGCCGACATCTCAACCA 5122

Db 216 AAGTGAAGACGCTCCGCGCCAGTCCCACTGCTGCCAGCCGACATCTCAACCA 157

QY 5123 TTTTGAAGCTGCTCAAGGCGTCCCACTGTTTGGGCGAGCAAGATTGAGACGCGCTC- 5181

Db 156 TTTTGAAGCTGCTCAAGGCGTCCCACTGTTTGGGCGAGCAAGATTGAGACGCGCTC 97

QY 5182 TGCATCTGTGACCGCATGCCCAAGATGCTCATCTCCGCTACAAAGAAACCTCAGC 5241

Db 96 TGCATCTGTGACCGCATGCCCAAGATGCTCATCTCCGCTACAAAGAAACCTCAGC 37

QY 5242 AATACTGCTATCCGGAAGATAGAGACTC 5273

Db 36 AATACTGCTATCCGGAAGATAGAGACTC 5

RESULT 15

LOCUS BUI38506 968 bp mRNA linear EST 25-NOV-2002

DEFINITION 603132657F1 CSEQCHL24 Gallus gallus cDNA clone CHEST11b15 5', mRNA sequence.

ACCESSION BUI38506

VERSION BUI38506.1 GI:25352689

SOURCE EST

ORGANISM Gallus gallus (chicken)

Gallus gallus

Archaeopteryx; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Eukaryota; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

1 (bases 1 to 968)

Boardman, P.E., Sant-Eguero, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Hickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J. A Comprehensive Collection of Chicken CDNAS

Curr. Biol. 12 (22), 1965-1969 (2002)

2235534

12445392

COMMENT Contact: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1OD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

1. 968

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="White Leghorn, Hisex"

/db\_xref="taxon:9031"

/clone="CHEST11b15"

/dev\_stage="16 day embryo"

/lab\_host="DH10B"

/clone\_id="CSEQCHL24"

/note="Organ: brain; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; Modification of pBluescript II KS(+): [Stratagene] vector to accommodate cDNA produced with the T-primed protocol (construction of uni-directionally cloned cDNA libraries from messenger RNA for improved 3' end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with NotI and EcoRI. Ligate in double stranded adaptor containing BspI and BamHI sites [5'ggcgcgtgcagcccgagatccgaaataaag] [5'aattcttttcgagaccggcgccagc]"

BASE COUNT 270 a 206 c 255 g 237 t

ORIGIN

Query Match 9.4%; Score 581.6; DB 13; Length 968;

Best Local Similarity 79.5%; Pred. No. 2.9e-127;

Matches 763; Conservative 0; Mismatches 189; Indels 8; Gaps 6;

QY 4393 AAGAGCCGACGACGACTTGCACCTGGAAGGTGATGAAGTCCCAAGATTAACAA 4452

Db 5 AAGAGCCGACGACGACTTGCACCTGGAAGGTGATGAAGTCCCAAGATTAACAA 64

QY 4453 CGAGACAGCAGGCTGGGACAGAGATACATTGCTCTGAGAGGATCAAAAGCTCAT 4512

Db 65 CGTGGCAACAGGCTGGGACAGAGATATATTGCTCTGAGAGGATCAAAAGCTCAT 124

QY 4513 TATGACATGAAGCCGAGAGCTGGACAGAGCCGCTGAGAAATTTAGCTGTGCTT 4572

Db 125 TATGACATGAAGCCGAGAGCTGGACAGAGCCGCTGAGAAATTTAGCTGTGCTT 184

QY 4573 CCCGAGGAGATATATATATATATATATATATATATATATATATATATATATAT 4632

Db 185 CCGATGGGAGATATATATATATATATATATATATATATATATATATATATATAT 244

QY 4633 AAGCAGATGCTCCATATATATATATATATATATATATATATATATATATATAT 4692

Db 245 AAGCAGATGCTCCATATATATATATATATATATATATATATATATATATATAT 304

QY 4693 GGGAGAACCTCTACTTGTAGCTCCAGCTTCCCTGACAAACGCGCTGTCACCGCC 4752

Db 305 GGTAGAACCTCTACTTGTAGCTCCAGCTTCCCTGACAAACGCGCTGTCACGCA 364

QY 4753 TTAGATCAGTGTGCGAGGTGGAGAGTTTCTAGGGAAGAGCAGAACTGATGTA 4812

Db 365 CTGGAATCATATATATATATATATATATATATATATATATATATATATATAT 424

QY 4813 CTGCTTGGAAATCTCCCTGCTGGAACCTGGAAGGTGATGACCGCTAGCATGACG 4872

Db 425 TTGCTTGGAAATCTCCCTGCTGGAACCTGGAAGGTGATGATATATATATATATAT 484

QY 4873 CTGCTTGGAAATCTCCCTGCTGGAACCTGGAAGGTGATGATATATATATATATAT 4932

Db 485 ATGCTTGGAAATCTCCCTGCTGGAACCTGGAAGGTGATGATATATATATATATAT 544

QY 4933 GTCTTGGAAATCTCCCTGCTGGAACCTGGAAGGTGATGATATATATATATATAT 4992

Db 545 GTATTAAGAACTCTTAAAGCATCTCCAGAGGTGATGATGATATATATATATATAT 604

QY 4993 ATCAAGACCTGGAAGAGCTATCTATATATATATATATATATATATATATATAT 5052

Db 605 ATCAAGACCTGGAAGAGCTATCTATATATATATATATATATATATATATATAT 664

QY 5053 GACGT-GAAGAAAGTGAAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5111

Db 665 GATTCGAAAGAGTGAAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 724

QY 5112 CTACCCCAACATTTTGAAGCTGTCAAGAGGCTGCCACTTGTTTGGGCGAGCAAGTTGA 5171

Db 725 CTACCCCAACATTTTGAAGCTGTCAAGAGGCTGCCACTTGTTTGGGCGAGGTTGA 784

QY 5172 GAAGGGGCTCG-CATCTGTCAGACCATGCGCCAGCAAGTGG-CATTTCTCGCTACAC 5229

Db 785 GAATGCTTTTGCATCTGTCAGACCATGCGCCAGCAAGTGG-CATTTCTCGCTACAC 844

QY 5230 GA-AAACCTCAGCAATATCTCATTCGGAAGAGAA--TAGAGCTTGAAGCCTGACG 5285

Db 845 GAGAGGCTTGAAGAGTCTGATCTATCCAGAAAGGAAATGAAACCTCAGAGCCTTGACG 904

QY 5286 CTGATTCACCTTCAACCAATTAACA-GTATCTCATTTGAAACCAATTAATTTACGAATTCG 5344

Db 905 TGTTCCTTTGACCACTTTACAGAAATCTTTTGGGAAACCATTAAGGCTATATGAATTTG 964

Search completed: November 15, 2003, 16:24:20

Job time : 7690.47 secs



Db 367 GACATCTATGCTATGAAAGTATGAGAGAGAGGCTTTATTGGCCAGAGCAGGTTTCA 426  
Qy 421 TTTTGTGGAAGAGCGGACATATTTATCTGGAAGCACAGCCCGTGGATCCCCCAATTA 480  
Db 427 TTTTGTGGAAGAGCGGACATATTTATCTGGAAGCACAGCCCGTGGATCCCCCAATTA 486  
Qy 481 CAGTATGCTCTTCAGAGCAAAAATACCTTTATCTGATCATGATATATCAGCTTGAGG 540  
Db 487 CAGTATGCTCTTCAGAGCAAAAATACCTTTATCTGATCATGATATATCAGCTTGAGG 546  
Qy 541 GACTTGCTGTACATTTTGAATGATATGAGAGCCAGTTAGTGAAGAACTGATACAGTTT 600  
Db 547 GACTTGCTGTACATTTTGAATGATATGAGAGCCAGTTAGTGAAGAACTGATACAGTTT 606  
Qy 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCCGTTTCACTGATGAGATAGTGCATGCA 660  
Db 607 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCCGTTTCACTGATGAGATAGTGCATGCA 666  
Qy 661 GACATCAAGCCTGAGAACATTTCTGTTGACCGCACAGACACATCAAGCTGATGATTTT 720  
Db 667 GACATCAAGCCTGAGAACATTTCTGTTGACCGCACAGACACATCAAGCTGATGATTTT 726  
Qy 721 GATATGCGCGGAAATATGATTTCAACAAGATGGAATGCAACTCCGATTGGGAGC 780  
Db 727 GATATGCGCGGAAATATGATTTCAACAAGATGGAATGCAACTCCGATTGGGAGC 786  
Qy 781 CCAGATTACATGAGCTCTGAAAGTGTGACTGTATGAAACGAGGATGAGAAAGCACCTAC 840  
Db 787 CCAGATTACATGAGCTCTGAAAGTGTGACTGTATGAAACGAGGATGAGAAAGCACCTAC 846  
Qy 841 GAGCTGAGCTGTGACTGTGTGCTAGTGGGCGGTGATGGCTTATGATGATATTTATGGA 900  
Db 847 GAGCTGAGCTGTGACTGTGTGCTAGTGGGCGGTGATGGCTTATGATGATATTTATGGA 906  
Qy 901 TCCCTCTGAGAGAGGAACTCTGCAAGAACCTTCAATTAACATATGAAATTTCCAGGG 960  
Db 907 TCCCTCTGAGAGAGGAACTCTGCAAGAACCTTCAATTAACATATGAAATTTCCAGGG 966  
Qy 961 TTTTGAATTTCCAGATGACCCCAAGATGAGCACTGATCTTGTATCTGATTCGAAAGC 1020  
Db 967 TTTTGAATTTCCAGATGACCCCAAGATGAGCACTGATCTTGTATCTGATTCGAAAGC 1026  
Qy 1021 TTTGTGTGCGGCGAGAAAGAGAGACTGAAATTTGAAGGTCTTGTGCTGCCATCTTTCTTC 1080  
Db 1027 TTTGTGTGCGGCGAGAAAGAGAGACTGAAATTTGAAGGTCTTGTGCTGCCATCTTTCTTC 1086  
Qy 1081 TCTAAATTTGACTGGAACAACATTTGTACTCTCTCCCTTGTGTTCCGACCTCAAG 1140  
Db 1087 TCTAAATTTGACTGGAACAACATTTGTACTCTCTCCCTTGTGTTCCGACCTCAAG 1146  
Qy 1141 TCTGACGATGACCTTCCAAATTTTATGAAACCAAGAAATTCGTGGGTTTCACTCTT 1200  
Db 1147 TCTGACGATGACCTTCCAAATTTTATGAAACCAAGAAATTCGTGGGTTTCACTCTT 1206  
Qy 1201 CCGTCCGAGCTGAGAGCCCTCAAGGCTTTCTGGGTGAAAGACTGCGCTTTGTGGGGTTTTCG 1260  
Db 1207 CCGTCCGAGCTGAGAGCCCTCAAGGCTTTCTGGGTGAAAGACTGCGCTTTGTGGGGTTTTCG 1266  
Qy 1261 TACAGCAAGGCACTGGGATTTCTTGTATGATCTGATCTGTTGTGTGGGTCTGGAATCC 1320  
Db 1267 TACAGCAAGGCACTGGGATTTCTTGTATGATCTGATCTGTTGTGTGGGTCTGGAATCC 1326  
Qy 1321 CCTGCAAGACTAGCTTCCATGAGAAAGAACTTCTCATCAAAAGCAAAAGCTTACAGAC 1380  
Db 1327 CCTGCAAGACTAGCTTCCATGAGAAAGAACTTCTCATCAAAAGCAAAAGCTTACAGAC 1386  
Qy 1381 TCTCAGGCAAGTGTCAAGAGT 1403  
Db 1387 TCTCAGGCAAGTGTCAAGAGT 1409

RESULT 2  
US-09-016-434-513

; Sequence 513, Application US/09016434  
; Patent No. 650938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREMITH  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 513:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 258 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BRAINMON01  
; CLONE: 2290031  
; US-09-016-434-513

Query Match 4.2%; Score 258; DB 4; Length 258;  
Best Local Similarity 100.0%; Pred. No. 1.3e-57;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 5204 GCAAGTGTGCTATTCGCTACAAAGAAACCTGAGCAATAGTCATCGGAAAGAGA 5263  
Db 1 GCAAGTGTGCTATTCGCTACAAAGAAACCTGAGCAATAGTCATCGGAAAGAGA 60  
Qy 5264 TAGAGACTCGAGAGCCCTGAGAGCTGATACCAATTACAGATATCCATTGGA 5323  
Db 61 TAGAGACTCGAGAGCCCTGAGAGCTGATACCAATTACAGATATCCATTGGA 120  
Qy 5324 CCAATTAATTTCAAGAAATTCAGATGAGAGCACTGACAGCTGAGAGAAATTCGATTAAGA 5383  
Db 121 CCAATTAATTTCAAGAAATTCAGATGAGAGCACTGACAGCTGAGAGAAATTCGATTAAGA 180  
Qy 5384 ATGACCAATTCCTGGACCTGCTGTGTTGCGGCTCTTCCAAAGCTTCCCTGTCTCAA 5443  
Db 181 ATGACCAATTCCTGGACCTGCTGTGTTGCGGCTCTTCCAAAGCTTCCCTGTCTCAA 240  
Qy 5444 TCGTCAAGTGAACGCG 5461  
Db 241 TCGTCAAGTGAACGCG 258

RESULT 3  
US-08-422-699A-12



Sequence 12, Application US/08422699A  
Patent No. 5955265  
GENERAL INFORMATION:  
APPLICANT: Brook, J. David  
APPLICANT: Housman, David E.  
APPLICANT: Shaw, Duncan J.  
APPLICANT: Harley, Helen G.  
APPLICANT: Johnson, Keith J.  
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC  
TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02713  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422,699A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/422,706  
FILING DATE:  
APPLICATION NUMBER: US 08/023,612  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/839,255  
FILING DATE: 20-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01545  
FILING DATE: 19-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00253  
FILING DATE: 05-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB9202485.0  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-5830A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2726 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-422-699A-12

Query Match 3.7%; Score 227.2; DB 2; Length 2726;  
Best Local Similarity 54.1%; Pred. No. 5,1e-49;  
Matches 535; Conservative 0; Mismatches 438; Indels 15; Gaps 3;

194 CTGCTCTGATGAAGATTAAAGCAGTGAAGCACTTTGTCCGAAGTATTCGACACCATAG 253  
243 CGGAAGCTGGCCAGAGCAAGTACGTGGCCGACTTTGTGAGTGGCGGAGCCCATCGTGG 302  
254 CTGAATTACAGAGCTCCAGCCTTGGCGAAAGAGCTTGAAGTCAAGAACTCTTGAAGTT 313  
303 TGAGGCTTAAGAGGCTCGACTGCGAAGAGGAGCACTTCGAGATTTGAAAGGATCGGAC 362  
314 GTGTCACTTGTGCTGAAGTGCAGTGTGTAAGAGAAAGCAACCGGGGACATCTATGTCTA 373

363 GCGGGGCGCTTACGCGAGGTAGCGGTAGTAAGATGAGACAGCGGCGCGGTATGCCA 422  
374 TGAAGATGATGAAGAGAGAGCTTTATTTGGCCAGAGAGCTTTCAATTTTTTGAGAG 433  
423 TGAAGATCATGAAGAGAGTGCAGATGCTGAAGAGGGGCGAGGTGTCTGCTCCGTGAG 482  
434 AGCGGAATATTTATCTCGAAGACAGAGCCCGTGTATCCCAATTTACGATGCTTTC 493  
483 AGAGGACCTGTGTGTGAATGAGGAGCCGGCGGTGATCAAGCAGCTGCACTTCGCTTCC 542  
494 AGAAGAAAATCACTTTATCTGTGATGAGAAATACGCTGGAGGGGAGCTTGTCTAC 553  
543 AGAATGAGACTTACCTGTACCTGTGTATGAGATTAAGTGGCGGGAGCTGTGACAC 602  
554 TTTGAAATGATATGATGAGAGCAGTATGATGAAACCTGATTAAGTTTACCTAGTAGC 613  
603 TGCTGAGCAAGTTTGGGAGCGGAGATTCGGCCGAGATGGCGGCTTCTACCTGGCGAGA 662  
614 TGATTTGGCTGTTCACAGCGCTTCATCTGATGAGATGATGATGATGATGATGATGATG 673  
663 TTGTATGCGCATTAAGCTGCGTGCACCGGCTTGGCTTACGTGACAGAGGACATCAACC 722  
674 AGAAGATTCGTTGACGCGCAGAGCAGAGCAGATGAGCTGATGATTTGATGCGCGCA 723  
723 AGAAGATTCGTTGACGCGCAGAGCAGATGAGCTGATGATTTGATGCGCGCA 782  
734 AAGTGAATTCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 793  
783 AGCTGGGGCAGATGAGAGCGTGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 842  
794 CTCTGAAGTGTGA--CTGTGATGAGACGAGGATGAGAAAGGACCTTACGCTGAGCT 850  
843 CCCCCGAGATCTGAGAGCTGTGGCGGTGGCGCTGGGACAGGCACTACGCGCCGAGT 902  
851 GTGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 910  
903 GTGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 962  
911 CAGAGGGAACCTCTGCGCAGAACCTTCATTAATTAATTAATTAATTAATTAATTAATTA 970  
963 AGCGGATTCAGAGGAGAGAGCTTATGAGAGATGCTTCACTTCAAGAGAGACCTCTCT 1022  
971 TTCCAGATGACCCCAAGTGAAGT--GACTTTTGTATCTGATTCAGAGCTTGTGT 1027  
1023 TCCCGCTGTGAGCAGAGGGGTCCCTGAGAGGCTCGAATCTTCACTTCAAGCGGTCTGT 1082  
1028 GCGGCGAGAAAGAGAGCTGAGAGTGAAGTCTTGTCTG-----CATCTTCT 1078  
1083 GTCCCGGAGAGACAGGCTGGCGCGGATGGAGAGGAGCACTTCGAGACATCCCTTCT 1142  
1079 TCTTAAATTTGACTGGAACAAATTCGTACTTCTTCCCTTCTGTTCCACCTCTA 1138  
1143 TCTTTGCTGCTGAGTGGATGTCTCCGGGACAGGCGTCCCTTTCACCGGATTTGG 1202  
1139 AGCTGAGATGACACCTTCAATTTTGA 1166  
1203 AAGGTGCCACCGACATGCAACTTGA 1230

RESULT 4  
US-08-422-706B-12  
Sequence 12, Application US/08422706B  
Patent No. 5977333  
GENERAL INFORMATION:  
APPLICANT: Brook, J. David  
APPLICANT: Housman, David E.  
APPLICANT: Shaw, Duncan J.  
APPLICANT: Harley, Helen G.  
APPLICANT: Johnson, Keith J.  
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC  
TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF  
NUMBER OF SEQUENCES: 14

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
;; STREET: Two Millicia Drive  
;; CITY: Lexington  
;; STATE: Massachusetts  
;; COUNTRY: US  
;; ZIP: 02173  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentln Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/422,706B  
;; FILING DATE: 14-APR-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/284,543  
;; FILING DATE: 08-AUG-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/023,612  
;; FILING DATE: 26-FEB-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/839,255  
;; FILING DATE: 20-FEB-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US93/01545  
;; FILING DATE: 19-FEB-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/GB93/00253  
;; FILING DATE: 05-FEB-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: GB9202485.0  
;; FILING DATE: 06-FEB-1992  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Granahan, Patricia  
;; REGISTRATION NUMBER: 32,227  
;; REFERENCE/DOCKET NUMBER: MIT-583042  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617-861-6240  
;; TELEFAX: 617-861-9540  
;; INFORMATION FOR SEQ ID NO: 12:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2726 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; US-08-422-706B-12

Query Match 3.7%; Score 227.2; DB 2; Length 2726;  
Best Local Similarity 54.1%; Pred. No. 5,1e-49;  
Matches 535; Conservative 0; Mismatches 438; Indels 15; Gaps 3;  
QY 194 CTGCTCTGATGAAGATTAAAGCAGCGAGCACTTGTCCGGAAGATTCCGACACCATAG 253  
DB 243 CCGAAGTGGCCCGAGACAAAGTACGTGGCCGACTTCTTGACATGGGCGAGCCCATCGTGG 302  
QY 254 CTGAAGTACAGAGGCTCCAGCCTTCCGCAAGAGACTTCGAAATCGAAGTCTTGAAGTT 313  
DB 303 TGAAGCTTAAGAGAGTCCGACATCGACAGAGGACGACTTGGAATTCGAAAGTGAATCGGAC 362  
QY 314 GTGTGCACTTTCTGTAAGTGCAGGTGGTAAAGAGAAAGCAACCGGGACATCTATGCTA 373  
DB 363 GGGGGGCGGTTCAGCAGAGGTAGCGGTAGTGAAGATGAAGACGAGCGGCGAGGTATGCGA 422  
QY 374 TGAAGTGTATGAAGAAGAGGCTTATTGGCCCGAGAGAGGTTTCATTTTGAAGAAG 433  
DB 423 TGAAGATCATGAACAAGTGGGACATGCTAAGAAGGGGCGAGGTGCTGCTTCGTTGAG 482  
QY 434 AGCGGAACATATATCTCGAAGACACAGCCGCTGATCCCGCAATTAAGTATGCTTTC 493  
DB 483 AGAGGAGCGTGTGTGAATGGGAGACCGGCGGTGATCAACGAGCTGCACTTCGCTTCC 542

QY 494 AGACAAAATACCTTTATCTGTATGTAATATACGCTGAGGGGACTTGTCTGAC 553  
DB 543 AGATGAGAACTACTGTAAGTGTATGAGAGTATACGTGGGCGGAGACCTGCTGACAC 602  
QY 554 TTTTGATATGATGAGAGCAAGTATGAAAACCTGATACATTTTACTACTGAGC 613  
DB 603 TGCTGACCAAGTTTGGGAGAGGATTCGGCGGAGATGGCGGCTTCTACTGGGGGAGA 662  
QY 614 TGATTTGGCTGTTCACAGCGTTTCATCTGATGGGATACGTGATGAGACATCAAGCTG 673  
DB 663 TTGTATAGCCATAGATCTCGGTGACACCGGCTTGGCTAGCTGACAGGACATCAACCG 722  
QY 674 AGAACATTTCTGTGACCGCACAGACACATCAAGCTGTGATTTTGGATTTGCGCGA 733  
DB 723 ACAACATCTCTGACCGCTGTGGCCACATCCGCTGGCCGACTTGGCTTCTGCTCA 782  
QY 734 AATGATTTCAACAAGATGTGATGCAATGCCAACTCCGATTTGGGACCCCAATTACAG 793  
DB 783 AGCTGGGGCAGATGGAACGGTGGCGGTGCTGTGGCTGTGGGACCCCAACTACCTGT 842  
QY 794 CTCCTGAAGTCTGA---CTGTGATGAACGGGAGTGAAGAAAGCACTACGCGCTGACT 850  
DB 843 CCCCAGATCTCTCAGGCTGTGGGCGGTGGGCTGGGACAGGACGCTACGGGCCGAGT 902  
QY 851 GTGACTGTGTCAAGTGGGCGCTGATGCTATGAGATGATTTATGGAGATCCCTTCTG 910  
DB 903 GTGACTGTGGGCGCTGGGTGATTCGCTTGAATGTTCTATGGGACAGCGCCCTTCT 962  
QY 911 CAGAGGGAACCTTGGCCGCAACCTTCAATAACATTATGATTTCCAGGGTTTTGAAT 970  
DB 963 ACGGGATTTCAAGCGCGAGACCTATAGGCAAGATGTCACATCAAGAGACCTCTCTC 1022  
QY 971 TTCCAGATGACCCCAAGTGAAGT---GACTTCTTGTATCTGATTTCAAGCTTGTGT 1027  
DB 1023 TGCCGCTGTGAGACGAAGGGTCCCTGAGAGGCTCGAGACTTATTCAGCGGTGCTGT 1082  
QY 1028 GCGGCGAAGAAAGAGACTGAAGTTGAAGTCTTTGCTG-----CATCTTTCT 1078  
DB 1083 GTCCCCCGAGACACAGCTGGGCGGGGTGAGACAGGAGTCCGACACATCCCTTCT 1142  
QY 1079 TCTCTAAATTTGACTGGAACAACATTCGTAACTCTCCCTCCCTCGTCCACCTCA 1138  
DB 1143 TCTTTGGCTGAGCTGGGATGATCTCCGGGACAGCGTCCCTTTTACACCGGATTTG 1202  
QY 1139 AGTCTGAGATGACACTCCCAATTTTGA 1166  
DB 1203 AAGTGCAACCGACATGCAACTTGA 1230

RESULT 5  
US-08-422-699A-8  
; Sequence 8, Application US/08422699A  
; Patent No. 5955265  
; GENERAL INFORMATION:  
; APPLICANT: Brook, J. David  
; APPLICANT: Housman, David E.  
; APPLICANT: Shaw, Duncan J.  
; APPLICANT: Harley, Helen G.  
; APPLICANT: Johnson, Keith J.  
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC  
; TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Millicia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422,699A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/422,706  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/023,612  
FILING DATE: 26-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/839,255  
FILING DATE: 20-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01545  
FILING DATE: 19-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00253  
FILING DATE: 05-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB9202485.0  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-5830A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2511 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1746  
US-08-422-699A-8

Query Match 3.6%; Score 220.4; DB 2; Length 2511;  
Best Local Similarity 54.6%; Pred. No. 2.9e-47;  
Matches 512; Conservative 0; Mismatches 411; Indels 15; Gaps 3;

244 GACACCATTAGCTGAGTACAGAGCTCCAGCCTTCGGCAAGAGACTTGGAAGTCAAGT 303  
43 GCCATCGTGTGAGGCTTAAGAGGCTCCGACTGCAGAGGAGCAGACTTGAAGTCTGAAG 102  
304 CTGTAGGTGTGTCACCTTGTCTGAGTGCAGGTGTAAGAGAAAGCAACCGGGGAC 363  
103 GTGATCGAAGCGGGCGCTTCAAGCGAGGTAGCGGTAGTGAAGTAAAGCAAGCGGGCAG 162  
364 ATCTATGCTATGAAGTATGATGAAGAAAGGCTTATTGCCCCAGACAGAGTTTCATTT 423  
163 GTGTATGCCATGAAGATCATGAACAAGTGGGACATGCTGAAGAGGGGAGAGTGTCTGCGC 222  
424 TTGAAGAGAGCGGCAATATATTTCTGAAGCACAAGCCCGTGTATCCCCCAATTACG 483  
223 TTCCGTGAGAGAGGAGCGTGTGTGATGGGAGCCGCGGTGATCACCAAGCTGCAC 282  
484 TATGCTTTCAGAGCAAAATCACCTTATCTGTCGTATGAATATCAACCTCGAGAGGAGAC 543  
283 TTGCGCTTCCAGAGTGAACCTACTGTAAGTCTGATGAGATTAAGTGGGCGGGAGC 342  
544 TTGCTGTACCTTTGAATGATATGAGAGCAGTTAGTGAAGAACTGATACAGTTTAC 603  
343 CTGCTGACACTGCTGAGCAAGTTTGGGAGCGGAGTTCCGGCCGAGATGCCCGCTTAC 402  
604 CTACTGAGCTGATTTTGGCTGTCACAGCTTATCTGATGAGGATACGTGATCGAGAC 663  
Db 403 CTGGGAGAGATTTGATGCGCATAGACTCGGTGACCGGCTTGGCTAGTGCACAGGGAC 462

664 ATCAAGCTTGAAATCTTCTGTTGACCCGACAGACACATCAAGCTGTGATTTTGA 723  
463 ATCAAAACCCGACAACTCTGTGAGCCGCTGTGGCCACATCGCTGGCCGACTTGGC 522  
724 TCTGGCGGGAATGATTTCAAAAGATGGAATGCCAACTCCCGATTGGAGACCCA 783  
523 TCTGTCTCAAGCTGGGCAATGGAAGGTGCGGTGCTGTGTGCTGTGGCACCCA 582  
784 GATTACATGACTCTGA--AGTGTGACTGTGATGAAGCGGATGGAAGAAAGCACCTAC 840  
583 GACTACCTCTCCCGGAGATCTCTGAGAGCTGTGGGCGGTGGCTGGACAGGACACTAC 642  
841 GGCCTGAGCTGTGACTGTGTGTCAAGTGGCGGTGATTCCTATGATGATTTATGAGAGA 900  
643 GGGCCCGAGGTGACTGTGTGGCGGTGGGTGTGATTCCTATGAAATTTCTATGGGACG 702  
901 TCCCGCTTGCAGAGGAACTCTGCGCAGAACCTTCAATTAATTTGATTTCCAGCGG 960  
703 ACGCCCTTCTACGGGATTTCAAGGCGGAGACTTATGCAAGATCTCCACTACAAAGAG 762  
961 TTTTGAAATTTCCAGATGACCCCAAGTGAAGT--GACTTCTGATCTGATTTCAA 1017  
763 CACCTCTCTGCGCTGCTGTGAGCAGAGGAGTCCCTGAGAGGCTGCAATCTTCACTCAG 822  
1018 AGCTTGTGTGTGCGCCAGAAAGAGACTGAAGTTGAAAGTCTTTGCTGC----- 1068  
823 CGGTGTGTGTGTCGCCCGGAGACAGCGGTGGCGGTGGAGAGGCACTTCCGAGCA 882  
1069 CATCCCTTCTCTTAAATGACTGGAACACATTTGTAATCTCTCCCTCCCTCGTT 1128  
883 CATCCCTTCTTTTGGCTCTGACTGAGATGATCTCTCCGGAACAGCGTCCCTTTTACA 942  
1129 CCACCTCTCAAGTCTGAGATGACACCTTCAATTTTGA 1166  
943 CCGATTTGAAAGGTGCCACCGACGACATGCAACTTGA 980

RESULT 6  
US-08-422-706B-8  
Sequence 8, Application US/08422706B  
Patent No. 5977333  
GENERAL INFORMATION:  
APPLICANT: Brook, J. David  
APPLICANT: Housman, David E.  
APPLICANT: Shaw, Duncan J.  
APPLICANT: Harley, Helen G.  
APPLICANT: Johnson, Keith J.  
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE MYOTONIC  
TITLE OF INVENTION: DYSTROPHY GENE AND USES THEREOF  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millilia Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02713  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/422,706B  
FILING DATE: 14-APR-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/284,543  
FILING DATE: 08-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/023,612  
FILING DATE: 26-FEB-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/839,255  
FILING DATE: 20-FEB-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01545  
FILING DATE: 19-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00253  
FILING DATE: 05-FEB-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB9202485.0  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: MIT-5830A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2511 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1746  
US-08-422-706B-8

Query Match 3.6%; Score 220.4; DB 2; Length 2511;  
Best Local Similarity 54.6%; Pred. No. 2.9e-47;  
Matches 512; Conservative 0; Mismatches 411; Indels 15; Gaps 3;

QY 244 GACACCAATAGCTGAGTTACAGAGCTCCAGCCTTGGCAAGAGCTTGAAGTCAAGT 303  
DB 43 GCCATCGGTGAGAGCTTAAAGAGTCCGACTGCAGAGGAGCACTTGAAGTCTGAAG 102  
QY 304 CTGTAGGTGAGTCACTTGTCTGAAGTGCAGGTGTAAGAGAAAGCAACCGGGGAC 363  
DB 103 GTGATCGAGCCGGGGCGCTTACGCGAGTACGTAAGTGAAGTAAACAGACGGGCGAG 162  
QY 364 ATCTATGCTATGAAGTATGAAGAAAGGCTTATTGGCCAGAGCAGCTTTCATT 423  
DB 163 GTGATGCCATGAGATCATGAACAAAGTGGACATGTAAGAGAGGGGAGAGTGTCTGC 222  
QY 424 TTTGAGAGAGCGGGAACATATATCTGAAGACACAGCCCGTGTATCCCAATTACG 483  
DB 223 TTCCGTGAGAGAGGAGCGTGTGTGAATGGGACCGCGGTGATCACAGAGCTGCAC 282  
QY 484 TATGCTTTGAGAACAAATCACTTATCTGTCTGATGAATATCAGCTGAGGGGAC 543  
DB 283 TTCCGCTTCCAGATGAGAACTACTCTGACTGTGATGAGTATTAAGTGGCGGGGAC 342  
QY 544 TTGCTGTCACTTTGAATGATGATGAGACCAATTAGTAAACCTGATACAGTTTAC 603  
DB 343 CTGCTGACACTGTGAGCAAGTTGGGAGCGGATTCGCGCGAGATGGCGCTTCTAC 402  
QY 604 CTAGCTGAGCTGATTTTGGCTGTCAAGCGTCTCATCTGATGGGATTAAGTCACTGAGAC 663  
DB 403 CTGGGGGAGATTGTATAGCCATGACTCGGTGGACCGGCTTGGTGAAGTGCACAGGGAC 462  
QY 664 ATCAAGCTGAGAACATTCTGTGAGCCGACAGAGACATCAAGCTGTGATTTTGA 723  
DB 463 ATCAAGCCGACACATCTGTGAGCCGCTGTGGCCACATTCGCTTGGCGGACTTGGC 522  
QY 724 TCTGCGGGAATGATTAACAAGATGTAATGCAAACTCCGATTGGGACCCCA 783  
DB 523 TCTTGCCCAAGCTCGGGGAGATGGAACGGTGGCTGTGAGGCTGTGGGACCCCA 582  
QY 784 GATTACATGGCTCTGGA---AGTGTGCTGTGATGAACGGGAGTGAAGGAGCACTAC 840

DB 583 GACTACTGTCCCGGAGATCTGCAGCTGTGGCGGCTGGGACAGCAGCTAC 642  
QY 841 GGCGTGAACGTGAGTGTGATGAGTGGCGGTGATGCTATGATGATTTATGGAGAG 900  
DB 643 GGGCCGAGTGTGAGTGTGGCGGTGATGCTATGATGATTTATGGAGAG 702  
QY 901 TCCGCTTCCGAGAGGAGACCTTGGCCAGAACCTTCAATATGATTAATTTCCAGCG 960  
DB 703 AGGCTTCTTACGGGAGATTCACCGCGGAGACCTATGCAAGATGCTCACAGAGAG 762  
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGCT---GACTTCTGTGATGATTCAA 1017  
DB 763 CACCTCTCTCGCCGTGTGAGAGAGAGGCTCCCTGAGAGAGCTCGAGACTTCAATCAG 822  
QY 1018 AGCTTTTGGCGGCGGAGAGAGAGATGTAAGTTGAAGTCTTGTCTG----- 1068  
DB 823 CGGTGTGTGTCCCGGAGACACGGTGGCGGGGTGAGAGAGGAGACTTCCGAGCA 882  
QY 1069 CATCCTTCTCTCTAAATGACTGGAACAACATTCGTAACCTCTCCCGCTTCTGTT 1128  
DB 883 CATCCCTTCTTCTTGGCTGAGCTGGAGTGTCTCCGGGACAGGCTGCCCTTTTACA 942  
QY 1129 CCACCTTCAAGCTGACGATGACACCTTCAATTTGA 1166  
DB 943 CCGAATTTGAAAGTGCACCGACACATGCAACTTCA 980

RESULT 7  
US-08-484-044-11  
Sequence 11, Application US/08484044  
GENERAL INFORMATION:  
APPLICANT: Caskey, C. T.  
APPLICANT: Fu, Ying-Hui  
APPLICANT: Friedman, David L.  
APPLICANT: Pizutti, Antonio  
APPLICANT: Fenwick, Raymond G.  
TITLE OF INVENTION: Diagnosis of Myotonic Muscular Dystrophy  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Fulbright & Jaworski, L.L.P.  
STREET: 1301 McKinney, Suite 5100  
CITY: Houston  
STATE: Texas  
COUNTRY: U.S.A.  
ZIP: 77010-3095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,044  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/019,940  
FILING DATE: 19-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul, Thomas D.  
REGISTRATION NUMBER: 32,714  
REFERENCE/DOCKET NUMBER: D-5443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713/651-5325  
TELEFAX: 713/651-5246  
TELEX: 762829  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3182 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

US-08-484-044-11

Query Match 3.6%; Score 220; DB 1; Length 3182;  
Best Local Similarity 54.6%; Pred. No. 4.2e-47;  
Matches 511; Conservative 0; Mismatches 410; Indels 15; Gaps 3;

QY 246 CACCATAGCTGAGATTACAGAGCTCCAGCCTTGGCAAGAGACTTGAATCAGAACTCT 305  
DB 743 CATCTGTGTAGAGCTTAAAGAGGCTCCAGCTGACAGAGGAGGAGCTTGAATCTGAAGGT 802  
QY 306 TGTAGGTGTGTCACTTGTCTGAAGTGAAGTGTGTGAAGAGAGAGAGAGAGAGAGAGAG 365  
DB 803 GATCGAGACGGGGGGCGTTCAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 862  
QY 366 CTATGCTATGAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425  
DB 863 GTATGCTATGAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 922  
QY 426 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485  
DB 923 CCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 982  
QY 486 TGCCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545  
DB 983 CCGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1042  
QY 546 GCTGTCACTTTTGAATGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605  
DB 1043 GCTGACACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1102  
QY 606 AGCTGAGCTGATTTTGGCTGTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 665  
DB 1103 GGGGAGAGATGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1162  
QY 666 CAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 725  
DB 1163 CAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1222  
QY 726 TGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 785  
DB 1223 TTGCCTCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1282  
QY 786 TTACATGAGCTCTCTGA---AGTGTGATGATGATGAACGGGAGAGAGAGAGAGAGAGAG 842  
DB 1283 CTACCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1342  
QY 843 CCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902  
DB 1343 GCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1402  
QY 903 CCGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 962  
DB 1403 GCGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1462  
QY 963 TTGAAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1019  
DB 1463 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1522  
QY 1020 CTGTGTGTGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1070  
DB 1523 GTTGTGTGTGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1582  
QY 1071 TCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1130  
DB 1583 TCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1642  
QY 1131 CACCTCTCAAGCTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1166  
DB 1643 GGATTTTGAAGGTGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1678

RESULT 8  
US-09-804-471A-3

; Sequence 3, Application US/09804471A  
; Patent No. 6479269  
; GENERAL INFORMATION:  
; APPLICANT: WEBSTER, Marion et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CLO01164  
; CURRENT APPLICATION NUMBER: US/09/804,471A  
; CURRENT FILING DATE: 2001-03-13  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 174493  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(174493)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-804-471A-3

Query Match 3.3%; Score 205; DB 4; Length 174493;  
Best Local Similarity 100.0%; Pred. No. 3.2e-42; Indels 0; Gaps 0;  
Matches 205; Conservative 0; Mismatches 0;

QY 753 GGTGAATGCCAACTCCGATTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 812  
DB 130289 GGTGAATGCCAACTCCGATTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130348  
QY 813 GATGAACGGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 872  
DB 130349 GATGAACGGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130408  
QY 873 GATTGCTATGAGATGATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 932  
DB 130409 GATTGCTATGAGATGATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130468  
QY 933 CTTCATTAACATTATGATTTCCAG 957  
DB 130469 CTTCATTAACATTATGATTTCCAG 130493

RESULT 9  
US-08-630-822A-61  
; Sequence 61, Application US/08630822A  
; Patent No. 5840695  
; GENERAL INFORMATION:  
; APPLICANT: FRANK, GLENN R.  
; APPLICANT: HUNTER, SHIRLEY WU  
; APPLICANT: WALLENFELDS, LYNDIA  
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS  
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS  
; NUMBER OF SEQUENCES: 107  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado  
; COUNTRY: U.S.A.  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; APPLICATION NUMBER: US/08/630,822A  
; FILING DATE: 11-APR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CONNELL, GARY J.  
; REGISTRATION NUMBER: 32,020

```

; REFERENCE/DOCKET NUMBER: 2618-17-C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2706 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5..2706
; US-08-630-822A-61

```

Query Match 3.2%; Score 195.6; DB 2; Length 2706;

Best Local Similarity 54.5%; Pred. No. 9e-41; Mismatches 384; Indels 27; Gaps 4;

```

Matches 493; Conservative 0;
QY 281 CAAGAGCTTCGAGTCAGAGTCTTGTAGTGTGTCACCTTGTCTGAAGTCAGAGTGC 340
DB 78 CAGATGATTTTAAATTAATAAGTTATTGTCGAGAGACATTTGGTGAAGTACAGTTAG 137
QY 341 TAAGAGAGAAAGCAACCGGGGACATCTATGAAAGTGAAGAAAGAGCTTTAT 400
DB 138 TCGACACAAATCAATCGACAAAGTTTGTATGAACGCTATCAAAATTTGAATGA 197
QY 401 TGGCCAGAGAGCAGTTTCAATTTTGAAGAGCGAATATTATCTCGAAGACAA 460
DB 198 TTAAGAGACCAAGCTGATTTTGTGGAGAAAGCTATATATGCTCATGCAAAAT 257
QY 461 GCCCGTATCCCAATTAACATATGCTTTCAGAGCAAAATCACTTATCTGTGCA 520
DB 258 CAGAAATGATGTGACATTTTGTCTTTCAGATCAAAATATCTTATATGTCGA 317
QY 521 TGAATATAGCTGAGAGGGAGCTTGTGCTGCTTGTGAATGATGAGACCAATTAG 580
DB 318 TGAATATATGCGGGGGGTGACTTGTGATGTTTATG-----TCCGATTAATGAATTC 371
QY 581 ATGAAACCTGATACAGTTTACCTTACCTGAGCTGATTTTGGCTGTTCACAGGCTCATC 640
DB 372 CAGAAAAATGGGCAATGTTCTATACATGAAAGTGTGCTAGACATTTGATACATTTCACT 431
QY 641 TGAATGATACGTGATGAGACATCAAGCTTGAAGAAATTCCTGTTGAACCGACAGAC 700
DB 432 CATGGGATTTGATCATCGATGATTAACCTGATATATGCTTCTAGCAAAATATGATC 491
QY 701 ACATCAAGCTGTGATTTTGTGATCTGCCCGGAAATGAAATTAACAAGATGATGAATG 760
DB 492 ATTTAAAGTTAGTGACTTTGGAACCTGTATGAAATGATACAGATGTTGGTACGTT 551
QY 761 CCAAACTCCCGATTGGGAGCCCGAGATTACATGAGCTCTGAAAGTGTGACTGTATGAACG 820
DB 552 CTATATATGCTGTGGAACGCGCTGATTAATTTCTCCGAAAGT-----TTTGCAATCCC 605
QY 821 GGGATGAAAAAGGACCTTACGCTTGAAGCTGTGATGCTGTGATGAGGCGGTATGCT 880
DB 606 AAGGTGTGAGAGAGTTTACGGTGTGGAATGCGATTTGGTGTGCGGAAATTTTGTGT 665
QY 881 ATGAGATGATTTATGGGAGATCCCTCTGCAAGGGAACCTGTGCAGAACCTTCAATA 940
DB 666 ATGAAATGTTATTTGAGAAACACCTTTTATGCAAGCAAGTTGTTGGTAATTAACGTA 725
QY 941 ACATTAATGATTTCCAGCGGTTTGTGAAATTTCCAGATGACCCCAAGTGAAGCACTGACT 1000
DB 726 AATATGATTCACAGAAACTCATTAATTTCTCCAGAAAGTGAATTAAGCAATATAG 785
QY 1001 TTCTTATCTGATTCAGAGCTTGTGTGGGCGCAGAAAG-----AGACTGAAGT 1051
DB 786 CCGGATCTTTGATACAGAGATTTTAAACAGACAGAAACACAGCGCTTATAGCGAATGAAG 845
QY 1052 TTGAAGGTCTTGCTCCATCTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1105

```

```

DB 846 TGAAGAAATTAAGACATCATTTTCTATTAATAATGATCATGACTTTTGACAAATTTAA 905
QY 1106 GTAACCTCTCCCTCCCTTGTGTTCCACCCCTCAAGTGTGAGCATGACACCTCCAAATTTG 1165
DB 906 GAGACTCTGCCCTCCTGATGAGTCCAGAGCTGAGTGTGATATGATACAGAACTTTG 965
QY 1166 ATGA 1169
DB 966 ATGA 969

```

# RESULT 10

US-09-005-069-61  
Sequence 61, Application US/09005069  
Patent No. 5932470

## GENERAL INFORMATION:

APPLICANT: FRANK, GLENN R.  
APPLICANT: HUNTER, SHIRLEY WU

TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS  
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver  
STATE: Colorado  
COUNTRY: U.S.A.

ZIP: 80203

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/005,069  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/630,822  
FILING DATE: 11-APR-1996

ATTORNEY/AGENT INFORMATION:  
NAME: CONNELL, GARY J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-17-C3

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 61:  
SEQUENCE CHARACTERISTICS:

LENGTH: 2706 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
FEATURE:

NAME/KEY: CDS  
LOCATION: 5..2706

US-09-005-069-61

Query Match 3.2%; Score 195.6; DB 2; Length 2706;

Best Local Similarity 54.5%; Pred. No. 9e-41; Mismatches 384; Indels 27; Gaps 4;

Matches 493; Conservative 0;

```

QY 281 CAAGAGCTTCGAGTCAGAGTCTTGTAGTGTGTCACCTTGTCTGAAGTCAGAGTGC 340
DB 78 CAGATGATTTTAAATTAATAAGTTATTGTCGAGAGACATTTGGTGAAGTACAGTTAG 137
QY 341 TAAGAGAGAAAGCAACCGGGGACATCTATGCTATGAAAGTGAAGAAAGAGCTTTAT 400
DB 138 TCGACACAAATCAATCGACAAAGTTTGTGCTATGAAAGCGCTATCAAAATTTGAATGA 197

```





Db 606 AAGTGTGAGAGAGTTACGTCGTGAATGCGATTGTCGTGGAAATTTTGT 665  
QY 881 ATGAGATGATTTATGAGAGATCCCTTCGACAGAGGAACTCTGCCAATCTTCAATA 940  
Db 666 ATGAATGTATTTGAGAGAAACACCTTTTATGACAGACGTTGGTGAACCTTACAGTA 725  
QY 941 ACATTATGATTTCCAGCGGTTTTGAAATTTCCAGATGACCCCAAGTAGCAGTACT 1000  
Db 726 AATTATGATCAGAGAAACTCATTTTCTTCAGAAAGTGAATTAAGCAATATG 785  
QY 1001 TTCTGATCTGATTTCAAGCTTGTGTGGGCCAGAAAG-----ACACTGAAT 1051  
Db 786 CCGATCTTTGATACAGGATTTTAAACAGACAGAACAGCGTTTAGCGAATATGA 845  
QY 1052 TTGAAGTCTTGTGCGCATCTTCTCTTAAATGACTGA-----ACAACATTC 1105  
Db 846 TCGAAGAAATTAAGACACATTCATTTTATTAATGATCAATGCACTTTTGCAATTTAA 905  
QY 1106 GTAATCTCTCCCTCCCTTCGTTCCACCCCTCAAGTCTGACATGACACTTCCAAATTTTG 1165  
Db 906 GAGACTCTGCCACCTGATGACAGAGCTGATGATGATGATATCAAGGAACTTTG 965  
QY 1166 ATGA 1169  
Db 966 ATGA 969

RESULT 12  
US-09-004-730A-20

; Sequence 20, Application US/09004730A  
; Patent No. 6485968  
; GENERAL INFORMATION:  
; APPLICANT: Weber, Eric  
; APPLICANT: Wu Hunter, Shirley  
; APPLICANT: Sim, Gek-Kee  
; APPLICANT: Frank, Glenn  
; APPLICANT: Wallenfelds, Lynda  
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS AND APPARATUS TO COLLECT SUC  
; FILE REFERENCE: 2618-17-C5-PUS-1  
; CURRENT APPLICATION NUMBER: US/09/004,730A  
; PRIOR FILING DATE: 1998-01-08  
; PRIOR APPLICATION NUMBER: PCT/97US/18669  
; NUMBER OF SEQ ID NOS: 150  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO: 20  
; LENGTH: 2706  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (5)..(2704)  
US-09-004-730A-20

## Query Match 3.2%; Score 195.6; DB 4; Length 2706;

Best Local Similarity 54.5%; Pred. No. 9e-41;  
Matches 493; Conservative 0; Mismatches 384; Indels 27; Gaps 4;

QY 281 CAAAGACTTCAGAGTCAGAGCTTGTAGTGTGTGCTACTTGTCTGAAGTCAGAGTGG 340  
Db 78 CAGATGATTTTAAATTAATAAGTATGTTGTCGAGACATTTGGTGAAGTACGTTG 137  
QY 341 TAAGAGAGAAAGCAACCGGAGACATCTATGCTATGTAAGTGAAGAAAGGCTTTAT 400  
Db 138 TCGACACAAATCAATGACAAAGTTTGTCTATGAAAGCCATCAAAATTTGAATGA 197  
QY 401 TGGCCAGAGAGAGAGTTTATTTTGAAGAGAGGGAACATATATCTCGAAGACAA 460  
Db 198 TTAAGAGACAGAGCTCTGATTTTGTGAGAGAAAGTATTAATGAGCTCATGCAAAAT 257  
QY 461 GCCCGTGAATCCCAATTAACAGTATGCTTTGAGAGCAAAATCACTTTATCTGGTCA 520  
Db 258 CAGATGATGATGATGATTAACATTTGCTTTTCAAGATCAAAATATCTTTATATGCTCA 317

QY 521 TGAATATACGCTTGAGAGGAGCTTGTCTGTCATCTTTGAATATGATATAGACCAAGTTAG 580  
Db 318 TGGATTAATATGCGGGGGGTGACTTGTGATGACTTATG-----TCCGATTAATGAATTC 371  
QY 581 ATGAATACCTGATACAGTTTATACCTAGCTGAGCTGATTTGGCTGTTCACAGCTTATC 640  
Db 372 CAGAAAAATGAGCAATGTTCTATACATGAAAGTGTGCTACACCTTGATACAAATTCAC 431  
QY 641 TGAATGAGATGATGATGAGACATCAAGCCAGAACCTTCGTTGCGCCAGACAGAC 700  
Db 432 CCATGGATTTGTATCATGATGATGATTTAACTGATATATGCTTTAGACAAATATAGTTC 491  
QY 701 ACATCAAGCTGTGATTTTGTGATCTGCGCGCAAAATGATTAATCAACAGATGTGATG 760  
Db 492 ATTTAAATGATGATGATGATTTGATGATGATGATGATGATGATGATGATGATGATG 551  
QY 761 CCAAACTCCGATTTGGAGACCCCAATTAATGATGATGATGATGATGATGATGATGATG 820  
Db 552 CTAAATATGCTGTTGGAACGCTGATTAATTAATTTCTCCGAAATG-----TTTGCAGTCCC 605  
QY 821 GGGATGAAAAAGCACTACGCTGACCTGATGATGATGATGATGATGATGATGATGATG 880  
Db 606 AAGTGTGAGAGAGTTTACGTCGTGAATGCGATTTGTTGTTGTTGTTGTTGTTGTTGTT 665  
QY 881 ATGAGATGATTTATGAGAGATCCCTTCGACAGAGGAACTCTGCCAATCTTCAATA 940  
Db 666 ATGAATGTATTTGAGAGAAACACCTTTTATGACAGACAGTTGTTGTTGTTGTTGTTGTT 725  
QY 941 ACATTATGATTTCCAGCGGTTTTGAAATTTCCAGATGACCCCAAGTAGCAGTACT 1000  
Db 726 AATTATGATCAGAGAAACTCATTTTCTTCAGAAAGTGAATTAAGCAATATG 785  
QY 1001 TTCTGATCTGATTTCAAGCTTGTGTGGGCCAGAAAG-----ACACTGAAT 1051  
Db 786 CCGATCTTTGATACAGGATTTTAAACAGACAGAACAGCGTTTAGCGAATATGA 845  
QY 1052 TTGAAGTCTTGTGCGCATCTTCTCTTAAATGACTGA-----ACAACATTC 1105  
Db 846 TCGAAGAAATTAAGACACATTCATTTTATTAATGATCAATGCACTTTTGCAATTTAA 905  
QY 1106 GTAATCTCTCCCTCCCTTCGTTCCACCCCTCAAGTCTGACATGACACTTCCAAATTTTG 1165  
Db 906 GAGACTCTGCCACCTGATGACAGAGCTGATGATGATGATATCAAGGAACTTTG 965  
QY 1166 ATGA 1169  
Db 966 ATGA 969

## RESULT 13

US-08-981-799A-20  
; Sequence 20, Application US/08981799A  
; Patent No. 6576238

; GENERAL INFORMATION:  
; APPLICANT: Weber, Eric  
; APPLICANT: Wu Hunter, Shirley  
; APPLICANT: Sim, Gek-Kee  
; APPLICANT: Frank, Glenn  
; APPLICANT: Wallenfelds, Lynda  
; TITLE OF INVENTION: "NOVEL ECTOPARASITE SALIVA PROTEINS AND APPARATUS TO COLLECT S  
; FILE REFERENCE: 2618-17-C5-PUS  
; CURRENT APPLICATION NUMBER: US/08/981,799A  
; PRIOR FILING DATE: 1998-08-27  
; PRIOR APPLICATION NUMBER: PCT/97/18669  
; NUMBER OF SEQ ID NOS: 150  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO: 20  
; LENGTH: 2706  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis

FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (5), (2704)  
US-08-981-799A-20

Query Match 3.2%; Score 195.6; DB 4; Length 2706;

Best Local Similarity 54.5%; Pred. No. 9e-41;  
Matches 493; Conservative 0; Mismatches 384; Indels 27; Gaps 4;

```

QY 281 CAAAGAGCTTCGAGTACAGAGTCTTGTAGTGTGTGTCATTGTGTCAGAGTGTG 340
DB 78 CAGATATTTTAATTAATTAAGATTATGTGTCAGAGAGATTGTGGAAGTACATTG 137
QY 341 TAAAGAGAAAGCAACCGGGGACATCTATGCTTGAAGTGAAGAAAGAGCTTTAT 400
DB 138 TGGACACAAATCACTGACACAAGTTTGTCTTGAACCGCCATCAAAATTTGAATGA 197
QY 401 TGGCCGAGGAGCAGTTCATTTTGTAGGAAGAGGGAACATATTATCTCGAAGACAA 460
DB 198 TTAAGAGACCAAGCTGCAATTTTGTGGGAAGACGTCAATTAATGCTCATGCAAAAT 257
QY 461 GCCCGTGATCCCAATTACAGTATGCTTTCAGAGACAAAATCACTTTATCTGTGTC 520
DB 258 CAGAAATGATGTACAAATTACATTTTGTTCAGATCAAAATATCTTTATATGTGTC 317
QY 521 TGGAAATACGCTGAGGGGACTGCTGCACTTTTGAATATATATGAGACCAATTAG 580
DB 318 TGAATATATGCCGGGGGGTACTGTGTGTATG-----TCCGATTATGAATTC 371
QY 581 ATGAAACCTGATACAGTTTAACTAGCTGAGCTGATTTTGGCTGTTTCAAGGCTCANC 640
DB 372 CAGAAAAATGGGAAAGTTCTATACATGAAGTGTGCTAGCACTTGATCAATTCACCT 431
QY 641 TGAATGAGATCGTGCATGACATCAAGCTTGAGAAATCTGTTGTCGTCAGACAGAC 700
DB 432 CCAATGGATTGTGACATCGTATGTTAACTGATATATGCTTACAGCAATATATGTC 491
QY 701 ACATCAAGCTGTGTGTTTGTGATCTGCCCGGAAATGAATTAACAAGATGTGAATG 760
DB 492 ATTTAAAGTTAGCTGCTTGTGAACCTGTATGAAATGATGAATGTTGTGTGTCGT 551
QY 761 CCAAACTCCCGATTGGAGACCCAGATTATACATGAGCTCTGAAGTGTGATGATAAG 820
DB 552 CTAATATGCTGTGTGAAGCCGTGATTAATTTCTCCGAAAGT-----TTTGCAATCCC 605
QY 821 GGGATGAAAGGCACTTACGCTGTGACCTGTGACCTGTGTGTCAGTGGCGGTGATTCCT 880
DB 606 AAGGTGTGAAGAGTTTACGCTGTGAAATGCGATTGGTGTGTCGTGGAAATTTTGTGT 665
QY 881 ATGAGATGATTAATGGAGATGCCCTGTGCAAGAGGAACCTGTGCCAAGACCTTGAAT 940
DB 666 ATGAATATGTTATTTGAGAAACACCTTTTATGCAAGACAGTTGTGTTGAACCTTACAGTA 725
QY 941 ACATTATGAATTTCCAGCGTTTGTGAAATTTCCAGATGACCCCAAGTGAAGAGTACCT 1000
DB 726 AATTATGATCAAGAAACCTCATTAACCTTTCTCCAGAAAGTGAATTAAGCCATATG 785
QY 1001 TTCTTATCTGATTTCAAGCTTGTGTGCGGCCAGAAAG-----AGACTGAAGT 1051
DB 786 CCGGATCTTTGATACAGAGATTTTAAACAGACAGAAACACAGCCTTTTAGGACAGAAATGAAG 845
QY 1052 TTGAAGGCTTTTGTGCTGCATCTTTCTCTCTTAAATTTGACCTG-----ACAATTC 1105
DB 846 TGGAGAAATTTAAGACATCTCATTTTCAATTAATATCAATGAGACTTTTGAACATTTAA 905
QY 1106 GTAATCTCTCTCCCTTCTGTTCCACCTCTGAAGTGTGACATGACACTCCAAATTTTG 1165
DB 906 GAGACTCTGCCCACTGTAGTGCAGAGCTGAGTGTATGATGATACAGGAACCTTTG 965
QY 1166 ATGA 1169
DB 966 ATGA 969

```

## RESULT 14

US-08-685-576-5  
; Sequence 5, Application US/0868576  
; Patent No. 5906819

## GENERAL INFORMATION:

APPLICANT: Kaibuchi, Kozo  
APPLICANT: Iwamatsu, Akihito  
APPLICANT: Nakano, Takeshi  
APPLICANT: Ito, Masaaki  
APPLICANT: Takahashi, No. 5906819uaki  
TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,576  
FILING DATE: 24-JUL-1996  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 7-325129  
FILING DATE: 20-NOV-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 8-17150  
FILING DATE: 05-JAN-1996  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 8-131206  
FILING DATE: 26-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16887/843  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136

## INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4363 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..4164  
US-08-685-576-5

Query Match 3.0%; Score 186; DB 2; Length 4363;

Best Local Similarity 52.1%; Pred. No. 3.8e-38;  
Matches 470; Conservative 0; Mismatches 420; Indels 12; Gaps 2;

```

QY 93 CAGAGGAACACACCTTTATGACTCAAGACAGATGTCCTTTCCGAGAAAGGAT 152
DB 78 CCAAGAGAGAGCTGAGAGCGCTATCCGAAACCTCGCTCCCAATCAAGCTGAGAGCTT 137
QY 153 ATTAGTGCCCTCTTTGTTCTCTTTGAAGAATGCACTCAGCTGCTGTGAAGATTAA 212
DB 138 GCTGATGCTTAATTTCTTGTGCTTGTATTTAGATTTTCCGCTTTGAGAAAAACAA 197
QY 213 GCACGTGACCACTTTGTGTCGGAAGATTTCCACACCAATAGCTGAATACAGAGCTCCA 272

```

198 GAACATGATATATTTCTTAATATGATGAGAAAATTTGAAAAATTCAAAAGGCTTACA 257  
273 GCGCTCGGCAAAAGACCTTGAAGTCAAGCTTGTAGTGTGCTCACTTGTGTAAGT 332  
258 GATGAGGACAGAGCTATGATGTTGTAAAAGTTATGGAAGAGGTGCTTTGGTGAAGT 317  
333 GCAGGTGTGTAAGAGAGAAAAGCAACCGGGGACATCTATGCTATGAAAAGTGATGAGAAGAA 392  
318 GCAGTTGGTGTGTCACAAAGGATCCAGAAAGTTATGCTATGAAAGCTTCTTAGTAAGTT 377  
393 GCGTTATATGGCCCGAGAGAGAGTTTCAATTTTATGAGAAGAGCGGACATATTTATCTCG 452  
378 TGAATATGATAAAAAGATCATGATTCGCTTTTATTTTGGAGAAAAGATATTATAGCCCTT 437  
453 AAGCAAGACCCCGTGGATCCCGCAATTTACAGATGCTTTCAGACAACAAAATCACCTTTA 512  
438 TGCCAAATAGCCCTGGGTGGTTCACGCTTTTATGCTTTCAGATGATAGTATCTGTA 497  
513 TCTGTCTATGGAATATCAAGCTTGAGGGGACTTGTCTCACTTTTGAATAGATATAGAGA 572  
498 CATGCTAATGAGATACATGCTGGTGGAGACCTTGTAAACCTTATGAGTAATTTATGATG- 556  
573 CCAATGATGATAAAACCTGATACAGTTTTCCTAGCTGAGCTGATTTGGCTGTACAG 632  
557 -----TGCCTGAAAAATGGGCCAAATTTTACCTGCTGAGTGTCTTCTGCTGAGTGC 611  
633 CGTTCATCTGATGGGATACGTGCATCGACATCAAGCTGAGAACATTTCTGTGACCG 692  
612 AATACACTCCATGGCTTTAATATCAGAGATGTTGAAGCTGACAAACATGCTCTTGGATTA 671  
693 CACAGGACATCAAGCTGGTGGATTTTGGATCTCCCGGAAAATGAATTCAAACAAGAT 752  
672 ACATGACATCTAAATATGACAGATTTTGGCACGTGATGAAAGATGAGAAACGGGAT 731  
753 GGTGATGCAAAATCCCGATTTGGAGCCCAATATACATGCTCTGCTGAGTCTGACTGT 812  
732 GGTATATGTGATATACAGAGCTTGGAAACCGGATTTATATACCTGAGGTTCTTAAATC 791  
813 GATGAAACGGGATGAGAAAAGCACCTACGGCTGAGCTGTGATGATGATGATGGGAGT 872  
792 ACAAGGGGATGAT-----GGTTTCTATGGCGGAAGATGATGATGATGATGATGAT 845  
873 GATTCCTATGATGATGATTTATGAGAGATCCCTTTCGACAGAGGAACCTCTGCCAAGAC 932  
846 TTTCTTTATGATGATGATGATGAGGAGATCTCCATTTTATGCGATTCATTTGTAGAAC 905  
933 CTTCAATACATTTATGAAATTTCCAGCGGTTTGAATTTCCAGATGACCCCAAGTGAG 992  
906 ATATGCAAAATTTATGATCATTAAGATTCATCTGTGTTTCCCTGAGATGCAAGAAATTC 965  
993 CA 994  
966 CA 967

RESULT 15  
US-09-016-434-1105  
; Sequence 1105, Application US/09016434  
; Patent No. 6500938  
; GENERAL INFORMATION:  
; APPLICANT: Janice Au-Young  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
; NUMBER OF SEQUENCES: 1490  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/016,434  
; FILING DATE: HEREMITH  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0002 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 1105:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4065 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: G1276900  
; US-09-016-434-1105

Query Match 2.8%; Score 175.4; DB 4; Length 4065;  
Best Local Similarity 51.1%; Pred. No. 2.1e-35;  
Matches 469; Conservative 0; Mismatches 436; Indels 12; Gaps 2;

145 GAAGGATATATGATGATCCCTCTTGTCTCTTTGAAGATGACATCAGCTCTCGATG 204  
82 GATGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 141  
205 AAGATTAAGACAGTGAAGCACTTTGTCGGAAGTATTCGACACCATAGCTGATTTACAG 264  
142 AAAAAAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 201  
265 GAGCTCCAGCTTGGGCAAGAGCTTGAAGTCAAGTCAAGTCTTGTGATGATGATGATGATGAT 324  
202 GATTTACGAATGAAAGCTGAAGATATGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 261  
325 GGTGAAGTCAAGTGTGAAGAGAGAAAGCAACCGGGGACATCTATGATGATGATGATGATGATGAT 384  
262 GGAAGATTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 321  
385 AAGAAAGAGCTTTATTTGAGCCAGAGAGGATTTTCAATTTTGTGAGAAAGCGGAACATA 444  
322 AGCAAAATTTGAATATGATTAAGAGATCTGATTTCTGTTTCTGGAAGAAAGGACATC 381  
445 TTATTCGAAGACCAAGCCCGTGTATCCCAATTAAGATGCTCTTCAAGACAAATAT 504  
382 ATGCTTTTGGCAACAGCTTGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 441  
505 CACCTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 564  
442 TATCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 501  
565 TATGAGACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 624  
502 TATGATG-----TGCCTGAAAAATGGGCAAGATCTATGCGAGAAAGTATGTTCTTGA 555  
625 GTTACACCGTTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 684  
556 TTGATGCAATTCATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 615  
685 GTTACCGCAAGGACATCAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 744

```

Db      616 CTGGATATAATCTGACATTGTAAGTATGACAGATTTTGTACTGTATGAAGATGAATAAG 675
Qy      745 AACAAATGGTGAATGCCAACTCCGATTTGGGACCCCAATTACATGGCTCTGAAAGTG 804
Db      676 GAAGGCATGGTACGATGTGATACAGCGGTTGGAAACCTGATTATATTTCCCTGAAAGT- 734
Qy      805 CTGACTGTATGAACGGGGATGGAAGGACCTACGCGCTGGACTGTGACTGGTGTCA 864
Db      735 -----ATTAAATCCCAAGGTGTTGATGTTATTATGGAAGAAATGTGACTGTGTG 789
Qy      865 GTGGGGGTGATTGCTATGATGATTTTATGGAGATCCCTTCGAGAGGGAACTCT 924
Db      790 GTTGGGGTATTATTATACGAAATGCTTGTAGTGATACACCTTTTATGAGATTCTTGG 849
Qy      925 GCCAGAACCTTCAATACATTATGAATTTCCAGCGGTTTGAATTTCCAGATGACCC 984
Db      850 GTTGGAACTTACAGTAAATATGAAACATAAAATTCACCTTCTCTGATGATTAAT 909
Qy      985 AAAGTGAGCAGTGACTTCTTGATCTGATTCAAAGCTGTGTGGCGGCCAGAAAGAGAGA 1044
Db      910 GACATATCAAAAAGCAAAAAACCTTATTGTGCTTCTTACTGACAGGGAAGTGAGG 969
Qy      1045 CTGAAGTTGAAGTCT 1061
Db      970 TTAGGGCGAAATGTGT 986

```

Search completed: November 15, 2003, 16:33:19  
 Job time : 276.099 secs

01-2175-2003

```

/standard_name="Single nucleotide polymorphism
replace(6065,G)
/*tag= b
/standard_name="Single nucleotide polymorphism"

```



PF 20-DEC-2001; 2001MO-US50497.  
XX  
XX 27-DEC-2000; 2000US-258335P.  
XX  
XX (LEXI-) LEXICON GENETICS INC.  
PA  
XX  
PI Yu X, Miranda M, Fiddle CJ;  
XX  
XX WPI; 2002-599796/64.  
DR P-FSDB; ABB81927.  
XX  
XX Novel polynucleotide encoding human proteins that are structurally  
PT similar to animal kinases, useful for drug screening, diagnosis, in  
PT gene therapy of disorders and diseases e.g. cancer and pharmacogenomic  
PT applications -  
PS  
PS Claim 1; Page 37-39; 50pp; English.  
XX  
XX The invention relates to a novel human protein that shares structural  
CC similarity with animal kinases, including serine-threonine kinases,  
CC particularly with animal rho-interacting kinases. The proteins of the  
CC invention have nootropic and cyostatic activity. The polynucleotides may  
CC have a use in gene therapy. The encoded novel polypeptides are useful for  
CC generating antibodies, as reagents in diagnostic assays, for identifying  
CC other cellular gene products related to NHP and as reagents in assays for  
CC screening for compounds that are useful in the treatment of mental,  
CC biological or medical disorders and diseases including cancer. The  
CC sequence encodes a novel human kinase of the invention.  
XX  
XX Sequence 6165 BP; 1735 A; 1550 C; 1679 G; 1201 T; 0 other;  
SQ  
Query Match 100.0%; Score 6165; DB 24; Length 6165;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTTGAAGTCAAAATATGAGCGCGGAATCTTTGATGCTGTGCTGAACCAT 60  
DB 1 ATGTTGAAGTCAAAATATGAGCGCGGAATCTTTGATGCTGTGCTGAACCAT 60  
QY 61 GCCAGCGGGCTCCAGGCTGAATCTGTTCTCCAGGGAACACCCCTTATGACTCA 120  
DB 61 GCCAGCGGGCTCCAGGCTGAATCTGTTCTCCAGGGAACACCCCTTATGACTCA 120  
QY 121 CACCAATGTCTCTCTTTCCCGAAGAGATATTAAGTCCCTCTTTGTTCTTTGAA 180  
DB 121 CACCAATGTCTCTCTTTCCCGAAGAGATATTAAGTCCCTCTTTGTTCTTTGAA 180  
QY 181 GAATGAGTCACTGCTGTGATGAAGATTAAAGCACTGAGCACTTGTCCGGAAGTAT 240  
DB 181 GAATGAGTCACTGCTGTGATGAAGATTAAAGCACTGAGCACTTGTGTCCGGAAGTAT 240  
QY 241 TCCGACACCAATGCTGATTAAGAGCTCCAGCTTCCGCAAGAGATTGCAAGTCA 300  
DB 241 TCCGACACCAATGCTGATTAAGAGCTCCAGCTTCCGCAAGAGATTGCAAGTCA 300  
QY 301 AGCTTGTAAGTGTGTGCTTGTGCTGAAGTCAAGTGTGAAGAGAAAGAACCGGG 360  
DB 301 AGCTTGTAAGTGTGTGCTTGTGCTGAAGTCAAGTGTGAAGAGAAAGAACCGGG 360  
QY 361 GACATCTATGCTATGAAGTGAAGAAAGAGCTTATTTGGCCCGAGAGAGTTTCA 420  
DB 361 GACATCTATGCTATGAAGTGAAGAAAGAGCTTATTTGGCCCGAGAGAGTTTCA 420  
QY 421 TTTTGTGAGGAAGAGGGAACATATTAATCTCGAAGCAAGCCCGTGTGATCCCAATTA 480  
DB 421 TTTTGTGAGGAAGAGGGAACATATTAATCTCGAAGCAAGCCCGTGTGATCCCAATTA 480  
QY 481 CAGTATGCTTGTGAGGAAGAAATCACTTTATCTGTGCTATGGAATATGAGGAGG 540  
DB 481 CAGTATGCTTGTGAGGAAGAAATCACTTTATCTGTGCTATGGAATATGAGGAGG 540  
QY 541 GACTTCTGTCACTTTTGAATAGATATGAGACCAAGTTAGATGAAAACTGATCAAGTTT 600  
DB 541 GACTTCTGTCACTTTTGAATAGATATGAGACCAAGTTAGATGAAAACTGATCAAGTTT 600

DB 541 GACTTCTGTCACTTTTGAATAGATATGAGACCAAGTTAGATGAAAACTGATCAAGTTT 600  
QY 601 TACCTAGTGAAGCTATTTTGGCTTTTCAAGGCTTCATCTGATGGGATACGTGATCGA 660  
DB 601 TACCTAGTGAAGCTATTTTGGCTTTTCAAGGCTTCATCTGATGGGATACGTGATCGA 660  
QY 661 GACATCAAGCTGAGAACATTTCTGTGACCGACAGAGACATCAAGCTGTGTGATTTT 720  
DB 661 GACATCAAGCTGAGAACATTTCTGTGACCGACAGAGACATCAAGCTGTGTGATTTT 720  
QY 721 GGAATCTGCGCGGAAATGAAATTCGAAATGCTGAAATTCGAAATTCGAAATTCGAAATTC 780  
DB 721 GGAATCTGCGCGGAAATGAAATTCGAAATGCTGAAATTCGAAATTCGAAATTCGAAATTC 780  
QY 781 CCAGATTACATGCTCTCTGAGAGCTGATGATGAACGGGATGGAAGGACCTTAC 840  
DB 781 CCAGATTACATGCTCTCTGAGAGCTGATGATGAACGGGATGGAAGGACCTTAC 840  
QY 841 GGCCTGAGCTGTGATGCTGTGCTGAGTGGGCTGATTCCTATGATGATTTATGGAGA 900  
DB 841 GGCCTGAGCTGTGATGCTGTGCTGAGTGGGCTGATTCCTATGATGATTTATGGAGA 900  
QY 901 TCCGCTTCCAGAGGGAACCTGCTGACAGAACTTCAATTAATTAATTAATTAATTAATTAAT 960  
DB 901 TCCGCTTCCAGAGGGAACCTGCTGACAGAACTTCAATTAATTAATTAATTAATTAATTAAT 960  
QY 961 TTTTGAATTTTCCAGATGACCCCAAGTGAAGTGAAGTCTTCTGATCTGATTTCAAGC 1020  
DB 961 TTTTGAATTTTCCAGATGACCCCAAGTGAAGTGAAGTCTTCTGATCTGATTTCAAGC 1020  
QY 1021 TTGTTGTGGGCGCAAGAGAGAGATGAAAGTCTTTGCTGCTGCTTCTTCTTCTTCTTCT 1080  
DB 1021 TTGTTGTGGGCGCAAGAGAGAGATGAAAGTCTTTGCTGCTGCTTCTTCTTCTTCTTCTTCT 1080  
QY 1081 TCTAAATGATCTGGAACCAATTCGTAATCTCCCTCCCTGCTGCTCCACCTCAAG 1140  
DB 1081 TCTAAATGATCTGGAACCAATTCGTAATCTCCCTCCCTGCTGCTCCACCTCAAG 1140  
QY 1141 TCTGAGATGACCTCCAAATTTTGTATGAACAGAGAAAGATTCGTGGTTTCACTCTCT 1200  
DB 1141 TCTGAGATGACCTCCAAATTTTGTATGAACAGAGAAAGATTCGTGGTTTCACTCTCT 1200  
QY 1201 CCGTGCAGCTGAGCCCTCAGGCTTCTCGGGTGAAGAACTCCGCTTTGTGGGGTTTTCG 1260  
DB 1201 CCGTGCAGCTGAGCCCTCAGGCTTCTCGGGTGAAGAACTCCGCTTTGTGGGGTTTTCG 1260  
QY 1261 TACAGAGGCACTGGGGAATTTCTGTGATGCTGAGTCTGTGTGTGTGTGTGTGTGTGTGTGT 1320  
DB 1261 TACAGAGGCACTGGGGAATTTCTGTGATGCTGAGTCTGTGTGTGTGTGTGTGTGTGTGTGT 1320  
QY 1321 CCTGCCAAGACTAGCTCAATGGAAGAAATTTCTGATCAAAAGCAAGAGCTTACAGAC 1380  
DB 1321 CCTGCCAAGACTAGCTCAATGGAAGAAATTTCTGATCAAAAGCAAGAGCTTACAGAC 1380  
QY 1381 TCTCAGAGCAAGTGTCAAGAGAGAGAGAGAAATGACCCGTTTACATCGGAAGTGTCA 1440  
DB 1381 TCTCAGAGCAAGTGTCAAGAGAGAGAGAGAAATGACCCGTTTACATCGGAAGTGTCA 1440  
QY 1441 GAGGTGAGGCTGTGCTTGTGCTGAGAGAGGTGAGGCTGAGAGCTTGTGAGCTGAGAGA 1500  
DB 1441 GAGGTGAGGCTGTGCTTGTGCTGAGAGAGGTGAGGCTTGTGAGCTTGTGAGCTGAGAGA 1500  
QY 1501 TCCCTCTGAGAGAGAGCTTGTGCTTACATCAAGATGCAAGTGTGCTTAAAGCGAAGT 1560  
DB 1501 TCCCTCTGAGAGAGAGCTTGTGCTTACATCAAGATGCAAGTGTGCTTAAAGCGAAGT 1560  
QY 1561 TTGGAGCAAGAGAGAGTGAAGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCAT 1620  
DB 1561 TTGGAGCAAGAGAGAGTGAAGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCAT 1620  
QY 1621 GATATCAGAGAGAGAGCGGAAAGCTCCAAAGAAATCAAAGAGAGAGATTCAGAGCTCAA 1680  
DB 1621 GATATCAGAGAGAGAGCGGAAAGCTCCAAAGAAATCAAAGAGAGAGATTCAGAGCTCAA 1680

```
QY 1681 GTGGAAGAAATGAGTTGATGATGAACTAGTTGGAAGAGATCTTGTCTCAGACGAAGA 1740
DB 1681 GTGGAAGAAATGAGTTGATGATGAACTAGTTGGAAGAGATCTTGTCTCAGACGAAGA 1740
QY 1741 CGAGTGAATCTTACGAATCTGAGTGAAGAGTCTCGGCTTGTGCTGAAGAAATTCAG 1800
DB 1741 CGAGTGAATCTTACGAATCTGAGTGAAGAGTCTCGGCTTGTGCTGAAGAAATTCAG 1800
QY 1801 CGGAAGGACACGAATGTCTACGATTAATCTGTTGAAGGCTTAAGATCAAGGAGGCTGA 1860
DB 1801 CGGAAGGACACGAATGTCTACGATTAATCTGTTGAAGGCTTAAGATCAAGGAGGCTGA 1860
QY 1861 GTGGAGATATGCGAACTGGAGAAAGATCAATGCTGAGCGACACTCAAAATTCAGAG 1920
DB 1861 GTGGAGATATGCGAACTGGAGAAAGATCAATGCTGAGCGACACTCAAAATTCAGAG 1920
QY 1921 CTCCAGAGAACTTGGAGAAAGCTGTAAAGCCAGACAGAGGCGCAGCTGCTGAG 1980
DB 1921 CTCCAGAGAACTTGGAGAAAGCTGTAAAGCCAGACAGAGGCGCAGCTGCTGAG 1980
QY 1981 AATATCCGCGCAGGCAAGAGAGCGAGCCGAGGGAGCTTGGAGAGCTGCAAGCCGAG 2040
DB 1981 AATATCCGCGCAGGCAAGAGAGCGAGCCGAGGGAGCTTGGAGAGCTGCAAGCCGAG 2040
QY 2041 GATTCCTTGAAGCATCAGAAAGAGCTGAGTGAAGCTGAAGGCTGAAGGCGCAGCTTCTG 2100
DB 2041 GATTCCTTGAAGCATCAGAAAGAGCTGAGTGAAGGCTGAAGGCGCAGCTTCTG 2100
QY 2101 GAGAACCAAGTAAAGAGCTAGAGACATGAGCGCTAGAGAAACAGACTAGAGATGAC 2160
DB 2101 GAGAACCAAGTAAAGAGCTAGAGACATGAGCGCTAGAGAAACAGACTAGAGATGAC 2160
QY 2161 ATCCAGACAAATCCCAACAGATCCAGACAGTGGCTGATTAATTTCTGAGCTGAGAG 2220
DB 2161 ATCCAGACAAATCCCAACAGATCCAGACAGTGGCTGATTAATTTCTGAGCTGAGAG 2220
QY 2221 AAACATCGAGAGGCGCAAGTCTCAGGCCAGACACCTAGAAAGTGCCCTGAACAGAAAG 2280
DB 2221 AAACATCGAGAGGCGCAAGTCTCAGGCCAGACACCTAGAAAGTGCCCTGAACAGAAAG 2280
QY 2281 CAGCACTATGAGAAAGATTAAAGTGTGGACATCAGATTAAGAAAGACCTGAGTAC 2340
DB 2281 CAGCACTATGAGAAAGATTAAAGTGTGGACATCAGATTAAGAAAGACCTGAGTAC 2340
QY 2341 AAGGAGACCTGGAGAACATGATGACAGACACAGAGGAGGCCCACTGAGAGGCGCAA 2400
DB 2341 AAGGAGACCTGGAGAACATGATGACAGACACAGAGGAGGCCCACTGAGAGGCGCAA 2400
QY 2401 ATTCTCAGCGAACAGAAAGGCGATGATCAATGCTATGATTCAGAGATCAGATCCCTGAA 2460
DB 2401 ATTCTCAGCGAACAGAAAGGCGATGATCAATGCTATGATTCAGAGATCAGATCCCTGAA 2460
QY 2461 CAGAGATTTGTGAATCTGTCTGAAGCCATTAACCTTCAGCAAAATACAGCTTTTACC 2520
DB 2461 CAGAGATTTGTGAATCTGTCTGAAGCCATTAACCTTCAGCAAAATACAGCTTTTACC 2520
QY 2521 CAAAGGAACTGAAGGCGCCAAAGAGATGATTTCTAAGCTCAGGCAACAGAAATTTAC 2580
DB 2521 CAAAGGAACTGAAGGCGCCAAAGAGATGATTTCTAAGCTCAGGCAACAGAAATTTAC 2580
QY 2581 CTGGAGACACAGGCTGGAGATTGGAGGCCAGAACGAAACCTGGAGGAGCGCTGAG 2640
DB 2581 CTGGAGACACAGGCTGGAGATTGGAGGCCAGAACGAAACCTGGAGGAGCGCTGAG 2640
QY 2641 AAGATCAGCCACAGACCAAGTGAAGATGAGTGGCTGAGAACTGAGACCAAGATTG 2700
DB 2641 AAGATCAGCCACAGACCAAGTGAAGATGAGTGGCTGAGAACTGAGACCAAGATTG 2700
QY 2701 CGGAGGTCAGTCTTAGAGCAGAGAGCAAGAACTGAGCTCAAGGCGCAGCTACAGAG 2760
DB 2701 CGGAGGTCAGTCTTAGAGCAGAGAGCAAGAACTGAGCTCAAGGCGCAGCTACAGAG 2760
QY 2761 CTTACAGCTTCCCTGAGAGGCGGAGATCAAGTTGACAGCCCTGAGGCTGACGAGCG 2820
DB 2761 CTTACAGCTTCCCTGAGAGGCGGAGATCAAGTTGACAGCCCTGAGGCTGACGAGCG 2820
QY 2821 GCCCTGAGAGGCTTCCGAGGCGAAAGAGAGCTGGAAGACCAACACAGAACT 2880
DB 2821 GCCCTGAGAGGCTTCCGAGGCGAAAGAGAGCTGGAAGACCAACACAGAACT 2880
QY 2881 GAAAGAGATTCAGAGCACTCAGCGACATAGATGAAATCAGGCGCAATTTGATGCT 2940
DB 2881 GAAAGAGATTCAGAGCACTCAGCGACATAGATGAAATCAGGCGCAATTTGATGCT 2940
QY 2941 CTTGCTAACAGCTGTACTGTATTCACAGACTGAGAGACAGCTAAACAGCTGACCGAG 3000
DB 2941 CTTGCTAACAGCTGTACTGTATTCACAGACTGAGAGACAGCTAAACAGCTGACCGAG 3000
QY 3001 GACAAAGCTGAATCAACAAACCAAACTTTCTTCTGTCACCACTGATGAGCTTCT 3060
DB 3001 GACAAAGCTGAATCAACAAACCAAACTTTCTTCTGTCACCACTGATGAGCTTCT 3060
QY 3061 GCGCGCAACGAGATGATACACTGCGAAGTGAAGTGAACCATCTCGCGCGGAGATC 3120
DB 3061 GCGCGCAACGAGATGATGATACACTGCGAAGTGAAGTGAACCATCTCGCGCGGAGATC 3120
QY 3121 ACGGACGAGAGATGACCTTACAGCCAGAAAGCAAGATGAGGCTCTGAAGACACG 3180
DB 3121 ACGGACGAGAGATGACCTTACAGCCAGAAAGCAAGATGAGGCTCTGAAGACACG 3180
QY 3181 TGCAACATGCTGAGAGAAACAGTCAATGATTTGAGAGGCTTAAAGATGAGCTGTAGAA 3240
DB 3181 TGCAACATGCTGAGAGAAACAGTCAATGATTTGAGAGGCTTAAAGATGAGCTGTAGAA 3240
QY 3241 AAAGGCGGAGTGGAGAGGCTTGGAGAGGCTCTCGGCTGATGAGAAATCCAGTTGAG 3300
DB 3241 AAAGGCGGAGTGGAGAGGCTTGGAGAGGCTCTCGGCTGATGAGAAATCCAGTTGAG 3300
QY 3301 TGTGGGTTGAGAGCTGACAGAGATGCTGAGACACCGAAGAACAGAGAGGCGAGAGCC 3360
DB 3301 TGTGGGTTGAGAGCTGACAGAGATGCTGAGACACCGAAGAACAGAGAGGCGAGAGCC 3360
QY 3361 GATCAGCGGATCAACGAGCTTCCGCAAGTGTGAGAGCTGAGAGTGAAGAGACCAAGCT 3420
DB 3361 GATCAGCGGATCAACGAGCTTCCGCAAGTGTGAGAGCTGAGAGTGAAGAGACCAAGCT 3420
QY 3421 GAGATTCCTGCTGACAGAGGCTCTCAAAAGAGCAAGAGCTGAAAGGCGGAGGCTCTCT 3480
DB 3421 GAGATTCCTGCTGACAGAGGCTCTCAAAAGAGCAAGAGCTGAAAGGCGGAGGCTCTCT 3480
QY 3481 GACAGCTCAATGACTGAGAGAGACATGCTATGCTTGAATGATGATGCCGAACTTA 3540
DB 3481 GACAGCTCAATGACTGAGAGAGACATGCTATGCTTGAATGATGATGCCGAACTTA 3540
QY 3541 CAGCAGAGGCTGAGACTGAAACGAGAGCTCAAAAGAGGCTTCTGGAAGAGCCAA 3600
DB 3541 CAGCAGAGGCTGAGACTGAAACGAGAGCTCAAAAGAGGCTTCTGGAAGAGCCAA 3600
QY 3601 TTACAGAGAGATGAGACTGAGAAATCAACTTTCCGCTGATGATCAAGGAGCTGAA 3660
DB 3601 TTACAGAGAGATGAGACTGAGAAATCAACTTTCCGCTGATGATCAAGGAGCTGAA 3660
QY 3661 GAAAGCTTGAATGAGGCTGATCTTGAAGACAGAAAGAACTGAGAGTATCAGCTG 3720
DB 3661 GAAAGCTTGAATGAGGCTGATCTTGAAGACAGAAAGAACTGAGAGTATCAGCTG 3720
QY 3721 GAAAGCTTGAATGAGGCTTGAATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3780
DB 3721 GAAAGCTTGAATGAGGCTTGAATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3780
QY 3781 CAAACCAACTCATTTGATTTTCTGCAAGCCAAATGAGCAACTGCTTAAGAAAGAAAG 3840
DB 3781 CAAACCAACTCATTTGATTTTCTGCAAGCCAAATGAGCAACTGCTTAAGAAAGAAAG 3840
QY 3841 GTTCTCTGAGTACATGAGCTGAAGCTGGCCCTGAGAGAGAAAGTCTGCTGCA 3900
```

```
Db      ||||| 3841 GTTCTTGAGTACATGAGCTGAACTG3CCTTGAGAGAGAAAGCTCGGTGCA 3900
Qy      ||||| 3901 GAGTAAAGAGAGCCCTTTCAGAAAGACCCGATCGAGTCCGCTCCGCCGGAGAAAGT 3960
Db      ||||| 3901 GAGTAAAGAGAGCCCTTTCAGAAAGACCCGATCGAGTCCGCTCCGCCGGAGAAAGT 3960
Qy      ||||| 3961 GCCCACCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 4020
Db      ||||| 3961 GCCCACCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 4020
Qy      ||||| 4021 ATGCGCATGTCCGCGCATCGTGGCGGTGCGGAGAGACCAAGCCAGTGCATGAGCTGCTG 4080
Db      ||||| 4021 ATGCGCATGTCCGCGCATCGTGGCGGTGCGGAGAGACCAAGCCAGTGCATGAGCTGCTG 4080
Qy      ||||| 4081 GCCCCGCGCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGAAATTAGTCGGCGTCTT 4140
Db      ||||| 4081 GCCCCGCGCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGAAATTAGTCGGCGTCTT 4140
Qy      ||||| 4141 AAGGAAACGATGACCAACAAATATTTCTACCGATTCAACGTAGAGACTGAACATGCGAGCC 4200
Db      ||||| 4141 AAGGAAACGATGACCAACAAATATTTCTACCGATTCAACGTAGAGACTGAACATGCGAGCC 4200
Qy      ||||| 4201 ACAAGGTGCTGTGTGTGATACCGTGCACTTTGGAGCGGAGATCCAAATGTCTC 4260
Db      ||||| 4201 ACAAGGTGCTGTGTGTGATACCGTGCACTTTGGAGCGGAGATCCAAATGTCTC 4260
Qy      ||||| 4261 GAAATGTCAGGTATGTGTCAACCCCAAGTGTCTCAGTGTCTTGCACGCACTTGCGCTTG 4320
Db      ||||| 4261 GAAATGTCAGGTATGTGTCAACCCCAAGTGTCTCAGTGTCTTGCACGCACTTGCGCTTG 4320
Qy      ||||| 4321 CCTGTGATATATGCGCACACTTTCACGAGGCTTCTGCGGTGACAAATGAATCTCCCA 4380
Db      ||||| 4321 CCTGTGATATATGCGCACACTTTCACGAGGCTTCTGCGGTGACAAATGAATCTCCCA 4380
Qy      ||||| 4381 GGTCTCAGACCAAGAGGCCAGACAGAGCTTGCACTTGAGAGGTGATGAGAGTGGCC 4440
Db      ||||| 4381 GGTCTCAGACCAAGAGGCCAGACAGAGCTTGCACTTGAGAGGTGATGAGAGTGGCC 4440
Qy      ||||| 4441 AAGAAATACAAACGAGAGACGAGAGCTGGGACAGAGATGATCTGCTGAGGAGTCA 4500
Db      ||||| 4441 AAGAAATACAAACGAGAGACGAGAGCTGGGACAGAGATGATCTGCTGAGGAGTCA 4500
Qy      ||||| 4501 AAAGTCTCATTTATGACATGAGGACAGAGAGCTGGACAGAGCCGGTGGAGAAATTT 4560
Db      ||||| 4501 AAAGTCTCATTTATGACATGAGGACAGAGAGCTGGACAGAGCCGGTGGAGAAATTT 4560
Qy      ||||| 4561 GAGCTGTGCTTCCCGACCGGAGATGTATCTATTATGATGTGCGGTGCTTCCGAACTC 4620
Db      ||||| 4561 GAGCTGTGCTTCCCGACCGGAGATGTATCTATTATGATGTGCGGTGCTTCCGAACTC 4620
Qy      ||||| 4621 GGAATAACAGCCAAAGCAAGATGCTCCATACATGATGAGATGGAATCTCACCCGCAAC 4680
Db      ||||| 4621 GGAATAACAGCCAAAGCAAGATGCTCCATACATGATGAGATGGAATCTCACCCGCAAC 4680
Qy      ||||| 4681 ACCTGTGCGCCGGGAGAACCTCTATCTTGATGCTCCAGCTTCCCTGACAAACAGCGC 4740
Db      ||||| 4681 ACCTGTGCGCCGGGAGAACCTCTATCTTGATGCTCCAGCTTCCCTGACAAACAGCGC 4740
Qy      ||||| 4741 TGGGTACACCGCTTGAATCAATGTGTGAGGTGGAGAGTTCTTGAAGAAAAAGCAGAA 4800
Db      ||||| 4741 TGGGTACACCGCTTGAATCAATGTGTGAGGTGGAGAGTTCTTGAAGAAAAAGCAGAA 4800
Qy      ||||| 4801 GCTGATGCTAAACGTGTTGAAAATCTCCTGCTGAAAATGAAAAGTGAATGACCGTCTAGAC 4860
Db      ||||| 4801 GCTGATGCTAAACGTGTTGAAAATCTCCTGCTGAAAATGAAAAGTGAATGACCGTCTAGAC 4860
Qy      ||||| 4861 ATGAATCTGACGCTGCTTCACTGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Db      ||||| 4861 ATGAATCTGACGCTGCTTCACTGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Qy      ||||| 4921 TACGCTGATGATGCTTGAATAAATCTCCCTAACCAATGTCCAGAAATTGAGACAGTCTTC 4980
Db      ||||| 4921 TACGCTGATGATGCTTGAATAAATCTCCCTAACCAATGTCCAGAAATTGAGACAGTCTTC 4980

Db      ||||| 4921 TACGCTGATGATGCTTGAATAAATCTCCCTAACCAATGTCCAGAAATTGAGACAGTCTTC 4980
Qy      ||||| 4981 CAATTTATTTATTAACAAGACCTGAGAAAGCTACTCATGTATAGAGAGAAAGCGGGCA 5040
Db      ||||| 4981 CAATTTATTTATTAACAAGACCTGAGAAAGCTACTCATGTATAGAGAGAAAGCGGGCA 5040
Qy      ||||| 5041 CTGTGTCTTGTGACGTGAAAGAAAGTGAACAGTCCCTGAGCCAGTCCCACTGCTGCC 5100
Db      ||||| 5041 CTGTGTCTTGTGACGTGAAAGAAAGTGAACAGTCCCTGAGCCAGTCCCACTGCTGCC 5100
Qy      ||||| 5101 CAGCCGACATCTCACCCCAATTTTGAAGCTGTCAAGGCTGCGCACTTGTGCGGCA 5160
Db      ||||| 5101 CAGCCGACATCTCACCCCAATTTTGAAGCTGTCAAGGCTGCGCACTTGTGCGGCA 5160
Qy      ||||| 5161 GGCAGATGAGAACGGGCTGTGATCTGTGAGCCATGCGCCAGAAAGTCATCTC 5220
Db      ||||| 5161 GGCAGATGAGAACGGGCTGTGATCTGTGAGCCATGCGCCAGAAAGTCATCTC 5220
Qy      ||||| 5221 CGCTACAGAAAACTCAGCAAAATCTGATCCGAAAGAGATAGAGCTCAGAGCC 5280
Db      ||||| 5221 CGCTACAGAAAACTCAGCAAAATCTGATCCGAAAGAGATAGAGCTCAGAGCC 5280
Qy      ||||| 5281 TGCAGCTGATATCACTTTCACAAATTCAGTATCTCATTTGGAACCAATAATTCAGAA 5340
Db      ||||| 5281 TGCAGCTGATATCACTTTCACAAATTCAGTATCTCATTTGGAACCAATAATTCAGAA 5340
Qy      ||||| 5341 ATCAGATGAAGCAGTACAGCTCGAGAAATCTGATTAAGATGACATTCCTTGGCA 5400
Db      ||||| 5341 ATCAGATGAAGCAGTACAGCTCGAGAAATCTGATTAAGATGACATTCCTTGGCA 5400
Qy      ||||| 5401 CCTGTGTGTGTCGCGCTCTTCCAAACAGCTTCCCTGTCTCAATCTGACAGTGAACAGC 5460
Db      ||||| 5401 CCTGTGTGTGTCGCGCTCTTCCAAACAGCTTCCCTGTCTCAATCTGACAGTGAACAGC 5460
Qy      ||||| 5461 GCAAGGACAGAGAGAGATCTTGTGTGTTTCCAGAAATTTGAAGTGTGTGTGATCT 5520
Db      ||||| 5461 GCAAGGACAGAGAGAGATCTTGTGTGTTTCCAGAAATTTGAAGTGTGTGTGATCT 5520
Qy      ||||| 5521 TACGAAAGAGTACAGCGGACAGACGATCTCAAGTGAAGTGCCTTACCTTGGCTTGGC 5580
Db      ||||| 5521 TACGAAAGAGTACAGCGGACAGACGATCTCAAGTGAAGTGCCTTACCTTGGCTTGGC 5580
Qy      ||||| 5581 TACAGAGAACCTATCTGTTGTGACCACTTCACTCACTCACTCACTCACTCACTCACTCA 5640
Db      ||||| 5581 TACAGAGAACCTATCTGTTGTGACCACTTCACTCACTCACTCACTCACTCACTCACTCA 5640
Qy      ||||| 5641 GCAAGCTCTCAGCAGAGACCCCTGCGGAGTACTTGAACATCCCGAACCCGCGCTAC 5700
Db      ||||| 5641 GCAAGCTCTCAGCAGAGACCCCTGCGGAGTACTTGAACATCCCGAACCCGCGCTAC 5700
Qy      ||||| 5701 CTGGGCTGCTGATTTCTCAGAGAGATTTACTTGGCGCTCATACAGAGTAATAATTA 5760
Db      ||||| 5701 CTGGGCTGCTGATTTCTCAGAGAGATTTACTTGGCGCTCATACAGAGTAATAATTA 5760
Qy      ||||| 5761 AAGGTCAATTTGCTGAAGGAAAATCTGTGAAGAGTCCGAGCATGGAACCAACCGGGGC 5820
Db      ||||| 5761 AAGGTCAATTTGCTGAAGGAAAATCTGTGAAGAGTCCGAGCATGGAACCAACCGGGGC 5820
Qy      ||||| 5821 CGGTCACTTCCGAGAGAGCCCAAGAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 5880
Db      ||||| 5821 CGGTCACTTCCGAGAGAGCCCAAGAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 5880
Qy      ||||| 5881 ACCAAGCGGTGAGCTTCCAGCCAGCGCGAGCCAGAGAGCCAGAGAGAGAGAGAGAGAG 5940
Db      ||||| 5881 ACCAAGCGGTGAGCTTCCAGCCAGCGCGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 5940
Qy      ||||| 5941 AGCAACCCCAACCGCTACCGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6000
Db      ||||| 5941 AGCAACCCCAACCGCTACCGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6000
Qy      ||||| 6001 CGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6060
Db      ||||| 6001 CGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6060
```

Qy 6061 CCCGAGAGGCTGTTTGAAGACAGACAGAGGCGGCTGCTGCGGAGACCGGTGAGACC 6120  
 Db 6061 CCCGAGAGGCTGTTTGAAGACAGACAGAGGCGGCTGCTGCGGAGACCGGTGAGACC 6120  
 Qy 6121 CCGCTGTCGAGGTGAACAGAGTCTGGGACCAAGTCTTCAATATA 6165  
 Db 6121 CCGCTGTCGAGGTGAACAGAGTCTGGGACCAAGTCTTCAATATA 6165

## RESULT 2

AAD38864  
 ID AAD38864 standard; cDNA; 6298 BP.

AC AAD38864;

DT 23-SEP-2002 (first entry)

DE Human kinase (PKIN)-21 cDNA.

Human; kinase; PKIN; cancer; immune system disorder; atherosclerosis;  
 acquired immune deficiency syndrome; AIDS; Addison's disease; allergy;  
 asthma; multiple sclerosis; psoriasis; arteriosclerosis; cirrhosis;  
 development; hepatitis; cardiovascular; hypertension; drug screening;  
 myocardial infarction; Goodpasture's syndrome; lipid disorder; growth;  
 fatty liver; Gaucher's disease; Niemann-Pick's disease; anorectic;  
 hypercholesterolemia; obesity; gene therapy; cytosolic; anti-hiv;  
 neuroprotective; hepatotropic; hypotensive; cardiac; nephrotropic;  
 hyperlipidaemia; enzyme; gene; ss.

OS Homo sapiens.

Key Location/Qualifiers  
 CDS 55..6219  
 /\*tag= a  
 /product= "Human kinase (PKIN)-21"

MO200233099-A2.

25-APR-2002.

20-OCT-2001; 2001MO-US47728.

20-OCT-2000; 2000US-242410P.  
 27-OCT-2000; 2000US-244068P.  
 03-NOV-2000; 2000US-245708P.  
 09-NOV-2000; 2000US-247672P.  
 16-NOV-2000; 2000US-249565P.  
 22-NOV-2000; 2000US-252730P.  
 01-DEC-2000; 2000US-250807P.

(INCY-) INCYTE GENOMICS INC.

PI Gururajan R, Baughn MR, Walla NK, Elliott VS, Xu Y, Arvinu C;  
 PI Yao MG, Ramkumar J, Ding L, Tang YT, Hafalia AJA, Nguyen DB;  
 PI Gandhi AR, Lu Y, Yue H, Burford N, Bandman O, Tribouley CM;  
 PI Lai PG, Reardon SA, Lu DAM, Borowsky ML, Thornton M, Swannaker A;  
 PI Thangavelu K, Khan FA, Ison CH;

WPI; 2002-454603/48.

P-PSDB; AAE24150.

New human kinase polypeptide, for diagnosing, preventing and treating  
 cancer, immune system disorders, growth and development disorders,  
 cardiovascular disorders and lipid disorders

Claim 5; Page 207-209; 210pp; English.

The invention relates human kinases (PKIN) and their corresponding  
 nucleic acid sequences. PKIN and its DNA are useful for diagnosing,  
 creating and preventing cancer, an immune system disorder (e.g.,  
 acquired immune deficiency syndrome (AIDS), Addison's disease, allergy,  
 asthma, atherosclerosis, multiple sclerosis, psoriasis), disorders

CC affecting growth and development (e.g., arteriosclerosis, cirrhosis,  
 CC hepatitis), cardiovascular disorder (e.g., hypertension, myocardial  
 CC infarction, Goodpasture's syndrome), and a lipid disorder (e.g., fatty  
 CC liver, Gaucher's disease, Niemann-Pick's disease, hypercholesterolemia,  
 CC hyperlipidaemia, obesity), and for assessing the effects of exogenous  
 CC compounds. Anti-PKIN antibody is useful in a diagnostic test for a  
 CC condition or a disease associated with the expression of PKIN in a  
 CC biological sample. A composition comprising PKIN or an agonist or  
 CC antagonist of PKIN is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional PKIN.  
 CC PKIN is useful in a number of drug screening techniques and to analyse  
 CC the proteome of a tissue or cell type. PKIN DNA is useful for creating  
 CC knockin humanised animals or transgenic animals to model human diseases,  
 CC and in somatic or germline gene therapy. The present sequence is human  
 CC PKIN cDNA.

Sequence 6298 BP; 1772 A; 1585 C; 1720 G; 1221 T; 0 other;

Query Match 99.9%; Score 6161.8; DB 24; Length 6298;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 6163; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGTTGAAGTTCAATATGAGCGGGAATCCTTGGATGCTGCTGAACCCATT 60  
 Db 55 ATGTTGAAGTTCAATATGAGCGGGAATCCTTGGATGCTGCTGAACCCATT 114  
 Qy 61 GCCAGCCGCGCTCCAGGCTGAATCTGTTCTCCAGGGAACCAACCTTTATGACTCA 120  
 Db 115 GCCAGCCGCGCTCCAGGCTGAATCTGTTCTCCAGGGAACCAACCTTTATGACTCA 174  
 Qy 121 CAGCAGATGTCCTCTCTTCCCGAAGGATATTAATGCTCTTGTCTTTGAA 180  
 Db 175 CAGCAGATGTCCTCTCTTCCCGAAGGATATTAATGCTCTTGTCTTTGAA 234  
 Qy 181 GAATGACGTCAGCTGCTGATGAGATTAAAGCAGTGAACCTTGTCCGGAAGAT 240  
 Db 235 GAATGACGTCAGCTGCTGATGAGATTAAAGCAGTGAACCTTGTCCGGAAGAT 294  
 Qy 241 TCCGACACCATAGCTGATTAACAGAGCTCCAGCTTCGGCAAGAGACTTGAATCAGA 300  
 Db 295 TCCGACACCATAGCTGATTAACAGAGCTCCAGCTTCGGCAAGAGACTTGAATCAGA 354  
 Qy 301 AGCTTGTAGTGTGTGCTCACTTTGCTGAAGTGCAGGTGTAAAGAGAAACACCGGG 360  
 Db 355 AGCTTGTAGTGTGTGCTCACTTTGCTGAAGTGCAGGTGTAAAGAGAAACACCGGG 414  
 Qy 361 GACATCATGCTATGAAGATGAAGAGAGGCTTTATGGCCAGGAGAGGTTTCA 420  
 Db 415 GACATCATGCTATGAAGATGAAGAGAGGCTTTATGGCCAGGAGAGGTTTCA 474  
 Qy 421 TTTTGAAGAGAGGAGACATATATCTCGAAGACAAAGCCGATGATCCCAATTA 480  
 Db 475 TTTTGAAGAGAGGAGACATATATCTCGAAGACAAAGCCGATGATCCCAATTA 534  
 Qy 481 CAGTATGCTTTAGAGACAAATTCACCTTATCTGCTCATGGAATATCAGCTGAGGG 540  
 Db 535 CAGTATGCTTTAGAGACAAATTCACCTTATCTGCTCATGGAATATCAGCTGAGGG 594  
 Qy 541 GACTTGTGCTACTTTGAATAGATATGAGGACCAAGTTGATGAATGAACCTGATACGTT 600  
 Db 595 GACTTGTGCTACTTTGAATAGATATGAGGACCAAGTTGATGAATGAACCTGATACGTT 654  
 Qy 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGGCTCATCTGATGAGATACGTGATCGA 660  
 Db 655 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGGCTCATCTGATGAGATACGTGATCGA 714  
 Qy 661 GACATGAAGCTGAGAACATCTCTGTCACCGACAGACACATCAAGCTGTGATTTT 720  
 Db 715 GACATGAAGCTGAGAACATCTCTGTCACCGACAGACACATCAAGCTGTGATTTT 774  
 Qy 721 GGATCGCGCGCAAAATGATTCAAACAAGATGTTGAATGCCAACTCCCGATGGGACC 780  
 Db 775 GGATCGCGCGCAAAATGATTCAAACAAGATGTTGAATGCCAACTCCCGATGGGACC 834

```
QY 781 CCAGATTACATGGCTCTCTGAAGTCTGAAGTCTGATGAACGGGGATGAAAAAGCACCTAC 840
Db 835 CCAGATTACATGGCTCTCTGAAGTCTGATGAACGGGGATGAAAAAGCACCTAC 894
QY 841 GGCCTGACCTGTGACTGTGTGTGACAGTGGGCGTGAATGCTTATGAGATGATTTATGGGAGA 900
Db 895 GGCCTGACCTGTGACTGTGTGTGACAGTGGGCGTGAATGCTTATGAGATGATTTATGGGAGA 954
QY 901 TCCCTCTTGCAGAGGAACTCTGCGCAGAACCTTCAATACATTAATGATTTCCACGG 960
Db 955 TCCCTCTTGCAGAGGAACTCTGCGCAGAACCTTCAATACATTAATGATTTCCACGG 1014
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTCTTCTTGATCTGATTCAGAC 1020
Db 1015 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTCTTCTTGATCTGATTCAGAC 1074
QY 1021 TTGTTGTGCGCCAGAAAGAGACGTAAGTTTGAAGTCTTGTGCGCATCTTCTTC 1080
Db 1075 TTGTTGTGCGCCAGAAAGAGACGTAAGTTTGAAGTCTTGTGCGCATCTTCTTC 1134
QY 1081 TCTAAATTTGACTGGAACAACTTGTGA CTCTCTCCCTCCCTGTTCCACCTCAAG 1140
Db 1135 TCTAAATTTGACTGGAACAACTTGTGA CTCTCTCCCTCCCTGTTCCACCTCAAG 1194
QY 1141 TCTGAGATGACACTCCAAATTTGATGAACAGAGAAATTTGCGGTTTCACTCT 1200
Db 1195 TCTGAGATGACACTCCAAATTTGATGAACAGAGAAATTTGCGGTTTCACTCT 1254
QY 1201 CCGTCCAGCTGAGCCCTCAGGCTTCTCGGGTGAAGAACTGCCGTTTGTGGGTTTTCG 1260
Db 1255 CCGTCCAGCTGAGCCCTCAGGCTTCTCGGGTGAAGAACTGCCGTTTGTGGGTTTTCG 1314
QY 1261 TACAGCAAGCACTGGGGAATTTGTGTGATCTGAAGTCTGTGTGCGGCTTGAACCTC 1320
Db 1315 TACAGCAAGCACTGGGGAATTTGTGTGATCTGAAGTCTGTGTGCGGCTTGAACCTC 1374
QY 1321 CCTGCAAGACTAGTCCATGGAAGAACTTCTCATCAAGAAAGCAAGAAAGCTTAAGAC 1380
Db 1375 CCTGCAAGACTAGTCCATGGAAGAACTTCTCATCAAGAAAGCAAGAAAGCTTAAGAC 1434
QY 1381 TCTCAGACAAAGTGTCAAGAATGAGAGAGAAAGACCCGGTTCATCGAGAGTGTCA 1440
Db 1435 TCTCAGACAAAGTGTCAAGAATGAGAGAGAAAGACCCGGTTCATCGAGAGTGTCA 1494
QY 1441 GAGGTGAGGCTGTCTTAAGTCAAGAGAGGTGAGCTGAAAGGCTTGAAGCTCAGAGA 1500
Db 1495 GAGGTGAGGCTGTCTTAAGTCAAGAGAGGTGAGCTGAAAGGCTTGAAGCTCAGAGA 1554
QY 1501 TCCCTCTGAGAGCAGGACCTTGTCTACCTACATCAAGAAAGCAAGTCTTAAGGAACT 1560
Db 1555 TCCCTCTGAGAGCAGGACCTTGTCTACCTACATCAAGAAAGCAAGTCTTAAGGAACT 1614
QY 1561 TTGAGCAAGCAAGATGAGAGGTCTCCAGAGAGATGA CAAAGCACTCAGACTTCTCAT 1620
Db 1615 TTGAGCAAGCAAGATGAGAGGTCTCCAGAGAGATGA CAAAGCACTCAGACTTCTCAT 1674
QY 1621 GATATCAAGAGCAGAGCCCGAAGTCCAAAGAAATCAAGAGCAGAGTACCAAGCTCAA 1680
Db 1675 GATATCAAGAGCAGAGCCCGAAGTCCAAAGAAATCAAGAGCAGAGTACCAAGCTCAA 1734
QY 1681 GTGGAAGAAATGAGGTTGATGATGAATCAGTTTGAAGAGATCTTGTCTCAGCAAGAA 1740
Db 1735 GTGGAAGAAATGAGGTTGATGATGAATCAGTTTGAAGAGATCTTGTCTCAGCAAGAA 1794
QY 1741 CGAGATGATCTCTA CGAATCTGAGCTGAGAGTCTCGGCTTGTGCTGAAGAAATTCAG 1800
Db 1795 CGAGATGATCTCTA CGAATCTGAGCTGAGAGTCTCGGCTTGTGCTGAAGAAATTCAG 1854
QY 1801 CGGAAGGAGAGATGTCAAGCTTAATGTTTGAAGCTTAAGATCAAGGAGAGCTGAA 1860
Db 1855 CGGAAGGAGAGATGTCAAGCTTAATGTTTGAAGCTTAAGATCAAGGAGAGCTGAA 1914
QY 1861 GTGGAGAAATATGCGAACTGAGAGAGATCAATGTCTAGACAGCTCAAAAATTCAGAG 1920
Db 1915 GTGGAGAAATATGCGAACTGAGAGAGATCAATGTCTAGACAGCTCAAAAATTCAGAG 1974
QY 1921 CTCCAAAGAAATCTGAGAGAGCTTAAAGCAGACAGAGAGCCACGAGCTCTGACAG 1980
Db 1975 CTCCAAAGAAATCTGAGAGAGCTTAAAGCAGACAGAGAGCCACGAGCTCTGACAG 2034
QY 1981 AATATCCGCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Db 2035 AATATCCGCAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
QY 2041 GATTTCTTGAAGGATCAAGAAAGCTGTGGAAGCTGAGAAAGCCGCCATTCCTCTG 2100
Db 2095 GATTTCTTGAAGGATCAAGAAAGAGCTGTGGAAGCTGAGAAAGCCGCCATTCCTCTG 2154
QY 2101 GAGAACAAAGTAAAGAGCTAGAGACATGAGCGGTAGAGAAACAGACTGAAGATGAC 2160
Db 2155 GAGAACAAAGTAAAGAGCTAGAGACATGAGCGGTAGAGAAACAGACTGAAGATGAC 2214
QY 2161 ATCCAGACAAATCCCAACAGATCCAGAGAGATGCTGAATTAATCTGAGAGCTCGAAGAG 2220
Db 2215 ATCCAGACAAATCCCAACAGATCCAGAGAGATGCTGAATTAATCTGAGAGCTCGAAGAG 2274
QY 2221 AAACATCGGAGGCCCAAGTCTAGCCCAAGCACTGAAGTCACTGAAAGCAGAAAGAG 2280
Db 2275 AAACATCGGAGGCCCAAGTCTAGCCCAAGCACTGAAGTCACTGAAAGCAGAAAGAG 2334
QY 2281 CAGCACTATGAGGAAAGATTAATGTTGAGCAATTCAGATTAAGAAAGACTGTGTC 2340
Db 2335 CAGCACTATGAGGAAAGATTAATGTTGAGCAATTCAGATTAAGAAAGACTGTGTC 2394
QY 2341 AAGAGACACTGAGAGACATGATGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db 2395 AAGAGACACTGAGAGACATGATGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2454
QY 2401 ATTCTCAGCGAACAGAGAGCGATGATCAATGCTATGAGATTCCAGATCAGATCCTGAA 2460
Db 2455 ATTCTCAGCGAACAGAGAGCGATGATCAATGCTATGAGATTCCAGATCAGATCCTGAA 2514
QY 2461 CAGAGGATTTGGAATCTGCTGAAGCCCAATTAATCTGAGAGCAATAGCACTTTTTCAC 2520
Db 2515 CAGAGGATTTGGAATCTGCTGAAGCCCAATTAATCTGAGAGCAATAGCACTTTTTCAC 2574
QY 2521 CAAAGGACATGAAGGCCCAAGAGAGATTTCTGAAGCTCAGGCAACAGAAATTTTAC 2580
Db 2575 CAAAGGACATGAAGGCCCAAGAGAGATTTCTGAAGCTCAGGCAACAGAAATTTTAC 2634
QY 2581 CTGAGACACAGGCTGAGAGATTGAGAGGCCCAAGACCGAAGAACTGAGAGAGCTGAGAG 2640
Db 2635 CTGAGACACAGGCTGAGAGATTGAGAGGCCCAAGACCGAAGAACTGAGAGAGCTGAGAG 2694
QY 2641 AAGATCAAGCCCAAGACCAAGTGA CAAAGATGCGCTGAGAACTGAGAGCAAGATTG 2700
Db 2695 AAGATCAAGCCCAAGACCAAGTGA CAAAGATGCGCTGAGAACTGAGAGCAAGATTG 2754
QY 2701 CGGAGGCTCAGTCTAGAGACAGAGAGAGAGAACTGAGAGCTCAAGCGCAGCTCAGAG 2760
Db 2755 CGGAGGCTCAGTCTAGAGACAGAGAGAGAGAACTGAGAGCTCAAGCGCAGCTCAGAG 2814
QY 2761 CTACAGCTCTCTCTGACAGAGCGGAGTCAAGTTTACAGGCTCTGACAGCTGACAGGCG 2820
Db 2815 CTACAGCTCTCTCTGACAGAGCGGAGTCAAGTTTACAGGCTCTGACAGCTGACAGGCG 2874
QY 2821 GCCCTGAGAGCCAGCTTCCGACAGAGCAAGACAGAGCTGGAAGAGACCAAGCGAGACT 2880
Db 2875 GCCCTGAGAGCCAGCTTCCGACAGAGCAAGAGCTGGAAGAGACCAAGCGAGACT 2934
QY 2881 GAAGAGAGATTCAGGCACTCAAGGCACTATGAGATGAATTCAGCGCAATTTGATGCT 2940
Db 2935 GAAGAGAGATTCAGGCACTCAAGGCACTATGAGATGAATTCAGCGCAATTTGATGCT 2994
QY 2941 CTTGTAAACAGCTGTACTGTATACAGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
```



Db 2995 CTTGTAACAGCTGTAAGTATGACAGACCTGGAGGAGCAGCTAAACACAGCTGACCGAG 3054  
Qy 3001 GACAACGCTGAATCTCAACAAACCAAACTTTCTTGTCCAAACAACTCGATGAGGCTTCT 3060  
Db 3055 GACAACGCTGAATCTCAACAAACCAAACTTTCTTGTCCAAACAACTCGATGAGGCTTCT 3114  
Qy 3061 GGGCGCAACGACGAGATTGTACAACTGGAGTAGAAGTAGACCACTCTCCGCGGAGATC 3120  
Db 3115 GGGCGCAACGACGAGATTGTACAACTGGAGTAGAAGTAGACCACTCTCCGCGGAGATC 3174  
Qy 3121 ACCGAAACGAGATGACGCTTACCAAGCAGAAAGCAAAAGATGAGGCTCTGAAGACCAAG 3180  
Db 3175 ACCGAAACGAGATGACGCTTACCAAGCAGAAAGCAAAAGATGAGGCTCTGAAGACCAAG 3234  
Qy 3181 TGCAACATGCTGAGAGAAACAGGTATGATTTGAGAGGCTTAAACGATGAGCTCTAGAA 3240  
Db 3235 TGCAACATGCTGAGAGAAACAGGTATGATTTGAGAGGCTTAAACGATGAGCTCTAGAA 3294  
Qy 3241 AAAGAGGAGGAGATGAGGAGGCTGAGAGAGGCTCTGGGTGATGAGAAATCCCACTTTGAG 3300  
Db 3295 AAAGAGGAGGAGATGAGGAGGCTGAGAGAGGCTCTGGGTGATGAGAAATCCCACTTTGAG 3354  
Qy 3301 TGTGAGGTTGAGAGCTGAGAGGATGCTGAGACCCAGAAACAGAGCAGGCGAGAGCC 3360  
Db 3355 TGTGAGGTTGAGAGCTGAGAGGATGCTGAGACCCAGAAACAGAGCAGGCGAGAGCC 3414  
Qy 3361 GATCAGGAGATCAACCGAGTCTCCGACAGTGTGAGAGCTGCGATGAAAGAGCAAGAGCT 3420  
Db 3415 GATCAGGAGATCAACCGAGTCTCCGACAGTGTGAGAGCTGCGATGAAAGAGCAAGAGCT 3474  
Qy 3421 GAGATTCGCTCTGACGACAGGCTCTCAAAAGACGAAAGCTGAAGGCGGAGAGCTTCT 3480  
Db 3475 GAGATTCGCTCTGACGACAGGCTCTCAAAAGACGAAAGCTGAAGGCGGAGAGCTTCT 3534  
Qy 3481 GACAAGCTCAATGACCTGAGAAAGACATGCTATGCTTGAATGAATGACCCGAGAGCTTA 3540  
Db 3535 GACAAGCTCAATGACCTGAGAAAGACATGCTATGCTTGAATGAATGACCCGAGAGCTTA 3594  
Qy 3541 CACGACAGCTGAGAGATGAAACGAGAGCTCAAAACAGGCTTCTGAGAGCAAGCCAAA 3600  
Db 3595 CACGACAGCTGAGAGATGAAACGAGAGCTCAAAACAGGCTTCTGAGAGCAAGCCAAA 3654  
Qy 3601 TTACAGAGCAGATGAGCCTGAGAAATCAATTTCCGTGATGATTCAGAGCTGCA 3660  
Db 3655 TTACAGAGCAGATGAGCCTGAGAAATCAATTTCCGTGATGATTCAGAGCTGCA 3714  
Qy 3661 GAAAGCTCTAGATCGGAGCTGATCTATGAGACAGAAAGAGTGAAGTGAAGTGAAGCTG 3720  
Db 3715 GAAAGCTCTAGATCGGAGCTGATCTATGAGACAGAAAGAGTGAAGTGAAGTGAAGCTG 3774  
Qy 3721 GAAAACATTCAGGTTCTCTATTTCTCATGAAAAGGTGAAAATGGAAGGCACTATTTCTCA 3780  
Db 3775 GAAAACATTCAGGTTCTCTATTTCTCATGAAAAGGTGAAAATGGAAGGCACTATTTCTCA 3834  
Qy 3781 CAAACCAAACTCATGATTTTCTGCAAGCCAAAATGAGCAACCTGCTTAAAAGAAAAG 3840  
Db 3835 CAAACCAAACTCATGATTTTCTGCAAGCCAAAATGAGCAACCTGCTTAAAAGAAAAG 3894  
Qy 3841 GTTCTCTGACGATCAATGAGCTGAAAGCTGAGGCTTGAAGAGAGAGAGTCTGCTGCA 3900  
Db 3895 GTTCTCTGACGATCAATGAGCTGAAAGCTGAGGCTTGAAGAGAGAGAGTCTGCTGCA 3954  
Qy 3901 GAGCTAGAGAGAGCCCTTTCAGAAAGACCCGATGAGAGCTCGGCTCCGCGAGAGAGCT 3960  
Db 3955 GAGCTAGAGAGAGCCCTTTCAGAAAGACCCGATGAGAGCTCGGCTCCGCGAGAGAGCT 4014  
Qy 3961 GCCACCGGAAAGCAACGAGCAACCCATCGACGCGAGCCGCGAGAGAGAGCT 4020  
Db 4015 GCCACCGGAAAGCAACGAGCAACCCATCGACGCGAGCCGCGAGAGAGAGCT 4074  
Qy 4021 ATGCGCATGTCGCGCATGCTGCGGTGCGAGAGCAACAGCCCAATGAGCTGAGCTGCTG 4080

Db 4075 ATGCGCATGTCGCGCATGCTGCGGTGCGAGAGCAACAGCCCAATGAGCTGAGCTGCTG 4134  
Qy 4081 GCCCGGCATTCAGCGCGAGAAAGAGTCTTCAATCTCAGAGAAATTTAGTCGAGCTT 4140  
Db 4135 GCCCGGCATTCAGCGCGAGAAAGAGTCTTCAATCTCAGAGAAATTTAGTCGAGCTT 4194  
Qy 4141 AAGGAAAGCATGACCAACATATTTCTTCAACGATTTCAACGATGAGATGAGAGAGCC 4200  
Db 4195 AAGGAAAGCATGACCAACATATTTCTTCAACGATTTCAACGATGAGATGAGAGAGCC 4254  
Qy 4201 ACAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4260  
Db 4255 ACAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4314  
Qy 4261 GAATGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4320  
Db 4315 GAATGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4374  
Qy 4321 CTGCTGAATATGTCACACACTTACCGAGGCTTCTGCGGTGACAAAATGAACTCCCA 4380  
Db 4375 CTGCTGAATATGTCACACACTTACCGAGGCTTCTGCGGTGACAAAATGAACTCCCA 4434  
Qy 4381 GGTCTCAGACCAAGAGCCGAGAGAGCTTGTGACCTGTGAAGGCTGATGAAGTGTGCC 4440  
Db 4435 GGTCTCAGACCAAGAGCCGAGAGAGCTTGTGACCTGTGAAGGCTGATGAAGTGTGCC 4494  
Qy 4441 AGGATATCAAAACGAGGACAGCAAGCTGGGACAGAAATGATCAATTTCTTGAAGGATCA 4500  
Db 4495 AGGATATCAAAACGAGGACAGCAAGCTGGGACAGAAATGATCAATTTCTTGAAGGATCA 4554  
Qy 4501 AAAGTCTCATTTATGCAATTAAGCCAGAGAAAGCTGACAGAGCGGTGAGAAATTT 4560  
Db 4555 AAAGTCTCATTTATGCAATTAAGCCAGAGAAAGCTGACAGAGCGGTGAGAAATTT 4614  
Qy 4561 GAGCTGTGCTTCCGACGGGAGATGATCTATTTCAATGAGTGTGCTGTGTGTGTGTGTGTGTGT 4620  
Db 4615 GAGCTGTGCTTCCGACGGGAGATGATCTATTTCAATGAGTGTGCTGTGTGTGTGTGTGTGTGT 4674  
Qy 4621 GCAAAATACAGCCAAAGAGATGCTCCATACATCTGAAATGGAATCTCACCCGACACCC 4680  
Db 4675 GCAAAATACAGCCAAAGAGATGCTCCATACATCTGAAATGGAATCTCACCCGACACCC 4734  
Qy 4681 ACCTGTGAGCGGAGAAACCTCTACTTGTGAGTCTCCAGCTTCCCTGACAAACAGCGC 4740  
Db 4735 ACCTGTGAGCGGAGAAACCTCTACTTGTGAGTCTCCAGCTTCCCTGACAAACAGCGC 4794  
Qy 4741 TGGGTACCGCTTAAATGATGATGCTGTGAGAGTGTGAGAGTGTGAGAGAGAGAGAGAG 4800  
Db 4795 TGGGTACCGCTTAAATGATGATGCTGTGAGAGTGTGAGAGTGTGAGAGAGAGAGAGAG 4854  
Qy 4801 GCTGATGCTAACTGCTTGAAGAACTCCGTGCTGAAACTGGAAGGTGATGACCGCTTGAAC 4860  
Db 4855 GCTGATGCTAACTGCTTGAAGAACTCCGTGCTGAAACTGGAAGGTGATGACCGCTTGAAC 4914  
Qy 4861 ATGAATGCAAGCTGCTCTTCACTGAGCAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920  
Db 4915 ATGAATGCAAGCTGCTCTTCACTGAGCAAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4974  
Qy 4921 TAGGCGCTGAATGTCTTGAAGAACTCCCTTAAACCAATGCTCCAGAGAAATGAGAGCTTTC 4980  
Db 4975 TAGGCGCTGAATGTCTTGAAGAACTCCCTTAAACCAATGCTCCAGAGAAATGAGAGCTTTC 5034  
Qy 4981 CAAATTTATATTAATCAAGAGCTGAGAGATCACTATGATGACAGAGAGAGAGAGAGAGAG 5040  
Db 5035 CAAATTTATATTAATCAAGAGCTGAGAGATCACTATGATGACAGAGAGAGAGAGAGAGAG 5094  
Qy 5041 CTGTGTCTTGTGAGAGTGAAGAAAGTGAACAGTCCCTGAGCCCAATCCCACTGCTGCC 5100  
Db 5095 CTGTGTCTTGTGAGAGTGAAGAAAGTGAACAGTCCCTGAGCCCAATCCCACTGCTGCC 5154  
Qy 5101 CAGCCGACATCTCACCCCAACTTTTGAAGCTGTCAAGAGCTGCTGCACTTTGTTGGGCA 5160  
Db 5155 CAGCCGACATCTCACCCCAACTTTTGAAGCTGTCAAGAGCTGCTGCACTTTGTTGGGCA 5214



```

QY 5161 GCGAAGATTGAGAACGGGCTCTGCACTGTGAGCCATGCCCAGCAAAAGTCGTCATCTC 5220
DB 5215 GCGAAGATTGAGAACGGGCTCTGCACTGTGAGCCATGCCCAGCAAAAGTCGTCATCTC 5274
QY 5221 CGCTACACGAAAACCTCAGCAATATCTGCATCCGAAAAGATAGACCTCGAGCCC 5280
DB 5275 CGCTACACGAAAACCTCAGCAATATCTGCATCCGAAAAGATAGACCTCGAGCCC 5334
QY 5281 TGGAGCTGTATCCACTTCACCAATATACAGTATCCTCATTTGGAACCAATAAATCTACGAA 5340
DB 5335 TGGAGCTGTATCCACTTCACCAATATACAGTATCCTCATTTGGAACCAATAAATCTACGAA 5394
QY 5341 ATGACATGAGAGAGTACAGCTCGAGAAATTCCTGATATAGATACCATTCCTTGGA 5400
DB 5395 ATGACATGAGAGAGTACAGCTCGAGAAATTCCTGATATAGATACCATTCCTTGGA 5454
QY 5401 CCTGCTGTGTTTGGCCGCTCTTCCAGAGCTTCCTGTCATTCGTGCGAGTGAACAGC 5460
DB 5455 CCTGCTGTGTTTGGCCGCTCTTCCAGAGCTTCCTGTCATTCGTGCGAGTGAACAGC 5514
QY 5461 GCGAGGCGAGAGAGAGTACTGCTGTGTTTCCAGAAATTTGGAAGTGTGCTGATTC 5520
DB 5515 GCGAGGCGAGAGAGAGTACTGCTGTGTTTCCAGAAATTTGGAAGTGTGCTGATTC 5574
QY 5521 TACGAGAGAGTACGCGCAGACAGATCTCAAGTGAAGTGCCTTACCTTTGCGCTTGGC 5580
DB 5575 TACGAGAGAGTACGCGCAGACAGATCTCAAGTGAAGTGCCTTACCTTTGCGCTTGGC 5634
QY 5581 TACGAGAGAGTACGCTGTTTGTGACCACTTCAACTCTCAAGTGAATTTGAGATTCAG 5640
DB 5635 TACGAGAGAGTACGCTGTTTGTGACCACTTCAACTCTCAAGTGAATTTGAGATTCAG 5694
QY 5641 GCAAGCTCTCAGAGAGAGCCCTGCGCAGCGTACCTGAGACATCCCGAACCAGCGCTAC 5700
DB 5695 GCAAGCTCTCAGAGAGAGCCCTGCGCAGCGTACCTGAGACATCCCGAACCAGCGCTAC 5754
QY 5701 CTGGGCTCTGCACTTTCTCAGAGAGGATTTACTTGGCGTCTTCATACAGATTAATTA 5760
DB 5755 CTGGGCTCTGCACTTTCTCAGAGAGGATTTACTTGGCGTCTTCATACAGATTAATTA 5814
QY 5761 AGGGTCACTTGTGCAAGAGGAAACCTGTGAAGAGTCCGGCACTGACACACCCGGGCG 5820
DB 5815 AGGGTCACTTGTGCAAGAGGAAACCTGTGAAGAGTCCGGCACTGACACACCCGGGCG 5874
QY 5821 CCGTCCACCTCCGCGAGCAGCCCAACAAGAGAGCCCAACGTAACAAGAGACATC 5880
DB 5875 CCGTCCACCTCCGCGAGCAGCCCAACAAGAGAGCCCAACGTAACAAGAGACATC 5934
QY 5881 ACCAAGCGGTGGCTCTCAGCCCAAGCGCGCCGAGAGCCCAAGCCGCGAGAGCCA 5940
DB 5935 ACCAAGCGGTGGCTCTCAGCCCAAGCGCGCCGAGAGCCCAAGCCGCGAGAGCCA 5994
QY 5941 AGCAACCCCAACCGCTACCGCGAGAGGCGGAGCCGAGTGTGCGAGGCAAGTCTCCGCG 6000
DB 5995 AGCAACCCCAACCGCTACCGCGAGAGGCGGAGCCGAGTGTGCGAGGCAAGTCTCCGCG 6054
QY 6001 CGCCCTCTGAGAGAGAGAGTCCCGCGCGAGTACTCAGCAAGCGAGAGAGCGGTCC 6060
DB 6055 CGCCCTCTGAGAGAGAGAGTCCCGCGCGAGTACTCAGCAAGCGAGAGAGCGGTCC 6114
QY 6061 CCGCGAGAGGCTGTTTGAAGACAGAGAGGCGCGGCTGCTCGGAGAGCGCTGAGAGACC 6120
DB 6115 CCGCGAGAGGCTGTTTGAAGACAGAGAGGCGCGGCTGCTCGGAGAGCGCTGAGAGACC 6174
QY 6121 CCGCTGTCCAGGAGTGAACAGAGTCTGGAGACAGTCTTCAATATTA 6165
DB 6175 CCGCTGTCCAGGAGTGAACAGAGTCTGGAGACAGTCTTCAATATTA 6219

```

RESULT 3  
 AAL55214  
 ID AAL55214 standard; DNA: 6165 BP.

```

XX AC AAL55214;
XX XX
DT 01-MAY-2003 (first entry)
XX DE
XX DE Human CRK encoding DNA sequence, SEQ ID No 1.
XX XX
XX KW Anorectic; hypotensive; cardiast; antilipemic; cerebroprotective;
XX KW Anticough; osteopathic; antiarthritic; cytosolic; antidepressant;
XX KW Immunomodulator; antitumor; tranquilizer; antiparkinsonian; nootropic;
XX KW Neuroprotective; antineoplastic; antidiabetic; analgesic;
XX KW human citron rho/rac-interacting kinase; enzyme; CRK; ameliorating;
XX KW obesity; comorbidities; cancer; anorexia; cachexia; bulimia;
XX KW central nervous system disorder; chronic obstructive pulmonary disease;
XX KW diabetes; pain; gene; ds.
XX XX
OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 1..6165
XX FT /tag= a
XX FT /product= "Human CRK protein"
XX XX
XX EN W02003004523-A1.
XX PD
XX PD 16-JAN-2003.
XX XX
XX PF 28-JUN-2002; 2002MO-BP07156.
XX XX
XX PR 02-JUL-2001; 2001US-301841P.
XX PR 11-DEC-2001; 2001US-338651P.
XX PR 25-APR-2002; 2002US-375014P.
XX XX
XX PA (FARB ) BAYER AG.
XX PI
XX PI Zhu Z;
XX XX
XX DR WPI; 2003-221576/21.
XX DR P-PSDB; AAO26959.
XX XX
XX PT New human citron rho/rac-interacting kinase (CRK) polypeptide and
XX PT polynucleotide, useful in preventing, ameliorating or treating diseases
XX PT associated with human CRK dysfunction, e.g. obesity, diabetes or
XX PT Alzheimer's disease -
XX XX
XX PS Example 1; Fig 1; 237pp; English.
XX XX
XX CC The invention relates to an isolated polynucleotide encoding a human
XX CC citron rho/rac-interacting kinase polypeptide. The isolated
XX CC polynucleotide comprises a 6165 or 8603 base pair sequence, given in the
XX CC specification. The human citron rho/rac-interacting kinase (CRK)
XX CC polypeptide and polynucleotide are useful in preventing, ameliorating, or
XX CC treating diseases associated with human CRK dysfunction such as obesity
XX CC and obesity-associated comorbidities (e.g. hypertension, coronary artery
XX CC disease, hyperlipidaemia, stroke, gout, osteoarthritis, some types of
XX CC cancer including endometrial, breast, prostate and colon cancer),
XX CC anorexia, cachexia, bulimia, central nervous system disorders (e.g. mood
XX CC disorders, anxiety disorders, Parkinson's disease or Alzheimer's
XX CC disease), chronic obstructive pulmonary disease or diabetes. These can
XX CC also be used to treat pain associated with the disorders. The human CRK
XX CC polypeptide is also useful in diagnostic assays or in genetic testing.
XX CC The expression vector or the reagent is useful in preparing a medicament
XX CC for modulating the activity of a human CRK in a disease, e.g. obesity, a
XX CC central nervous system disorder, or chronic obstructive pulmonary
XX CC disease. The fusion protein is useful for generating antibodies against a
XX CC CRK polypeptide and for use in various assay systems. The methods are
XX CC useful in producing and detecting the polynucleotide and polypeptide and
XX CC in screening for agents that modulate the activity of the human CRK
XX CC polypeptide. This polynucleotide sequence represents a DNA sequence
XX CC encoding a human CRK protein of the invention.
XX XX
XX SO Sequence 6165 BP; 1735 A; 1549 C; 1680 G; 1201 T; 0 other;

```

Query Match 99.9%; Score 6160.2; DB 25; Length 6165;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6162; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTTGAAGTCAAAATATGAGCGCGGAATCTTTGATGCTGCTGCTGAACCCATT 60  
DB 1 ATGTTGAAGTCAAAATATGAGCGCGGAATCTTTGATGCTGCTGCTGAACCCATT 60  
QY 61 GCCAGCGGCGCTCCAGGCTGAATCTGTTCTCCAGGGAACACCCCTTATAGATCA 120  
DB 61 GCCAGCGGCGCTCCAGGCTGAATCTGTTCTCCAGGGAACACCCCTTATAGATCA 120  
QY 121 CAGCAGATGCTCTCTCTTTCCCGAAGAGGATATAGATGCTCTTTGTTCTTTGAA 180  
DB 121 CAGCAGATGCTCTCTCTTTCCCGAAGAGGATATAGATGCTCTTTGTTCTTTGAA 180  
QY 181 GAATGCACTGACCTGCTCTGATGAATTAAGCAGTGAACAATTGTCGGAAGTAT 240  
DB 181 GAATGCACTGACCTGCTCTGATGAATTAAGCAGTGAACAATTGTCGGAAGTAT 240  
QY 241 TCCGACACATAGCTGATTAAGAGCTCCAGGCTCCGCAAGAGACTTCGAAGTCA 300  
DB 241 TCCGACACATAGCTGATTAAGAGCTCCAGGCTCCGCAAGAGACTTCGAAGTCA 300  
QY 301 AGCTTGTAGGTTGTGCTCACTTTGCTGAAGTCAAGTGTAAAGAGAAAGCAACG 360  
DB 301 AGCTTGTAGGTTGTGCTCACTTTGCTGAAGTCAAGTGTAAAGAGAAAGCAACG 360  
QY 361 GACATCTATGCTATGAAGTGAATGAAGAGGCTTTATTTGGCCCGAGAGGTTCA 420  
DB 361 GACATCTATGCTATGAAGTGAATGAAGAGGCTTTATTTGGCCCGAGAGGTTCA 420  
QY 421 TTTTGTAGAGAGCGGAACATTTATCTCGAAGCAAAAGCCGTGATCCCCAATTA 480  
DB 421 TTTTGTAGAGAGCGGAACATTTATCTCGAAGCAAAAGCCGTGATCCCCAATTA 480  
QY 481 CAGATGCTTTGAGAGCAAAATCACCCTTATCTGATGCAATGAATTCAGCTGAGAG 540  
DB 481 CAGATGCTTTGAGAGCAAAATCACCCTTATCTGATGCAATGAATTCAGCTGAGAG 540  
QY 541 GACTTGTGCTCACTTTGATATGAGAGCAAGTATGATGAAACCTGATATCACTTT 600  
DB 541 GACTTGTGCTCACTTTGATATGAGAGCAAGTATGATGAAACCTGATATCACTTT 600  
QY 601 TACTTACCTGAGCTGATTTTGTGCTTCAACGCTTATCTGATGGAATCGTCATCA 660  
DB 601 TACTTACCTGAGCTGATTTTGTGCTTCAACGCTTATCTGATGGAATCGTCATCA 660  
QY 661 GACATCAAGCTGAGAACTTCTGTTGACCGACAGAGACATCAAGCTGTGATTTT 720  
DB 661 GACATCAAGCTGAGAACTTCTGTTGACCGACAGAGACATCAAGCTGTGATTTT 720  
QY 721 GGATCTGCGCGGAAATGAAATCAACAGATGCTGATCAAACTCCGATTTGGAGC 780  
DB 721 GGATCTGCGCGGAAATGAAATCAACAGATGCTGATCAAACTCCGATTTGGAGC 780  
QY 781 CCAATTTACATGCTCCGTAAGTCTGATGATGAACGAGAGTGAAGAAAGCACTTAC 840  
DB 781 CCAATTTACATGCTCCGTAAGTCTGATGATGAACGAGAGTGAAGAAAGCACTTAC 840  
QY 841 GGCTGGAATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 900  
DB 841 GGCTGGAATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 900  
QY 901 TCCGCTTCCGAGAGGAGAACTCCGCAAGACCTTCAATTAAGATTTTCCAGCG 960  
DB 901 TCCGCTTCCGAGAGGAGAACTCCGCAAGACCTTCAATTAAGATTTTCCAGCG 960  
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020  
DB 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020  
QY 1021 TTGTTGTGCGGCGAGAAAGAGACTGAAGTTTGAAGTCTTGTGCTCATCTTTCTTC 1080

DB 1021 TTGTTGTGCGGCGAGAAAGAGACTGAAGTTTGAAGTCTTGTGCTCATCTTTCTTC 1080  
QY 1081 TCTTAAATTTGACTGGAACAACCTTGTACTCTCTCCCTCTTGTTCACCTCAAG 1140  
DB 1081 TCTTAAATTTGACTGGAACAACCTTGTACTCTCTCCCTCTTGTTCACCTCAAG 1140  
QY 1141 TCTGAGATGACACCTCCCAATTTTATGAACAGAGAAATGCTGAGGTTTCACTCT 1200  
DB 1141 TCTGAGATGACACCTCCCAATTTTATGAACAGAGAAATGCTGAGGTTTCACTCT 1200  
QY 1201 CCGTGCACGCTAGGCTCCAGGCTTCTCGAGTGAAGAACTGCGGTTTGTGAGGTTTTCG 1260  
DB 1201 CCGTGCACGCTAGGCTCCAGGCTTCTCGAGTGAAGAACTGCGGTTTGTGAGGTTTTCG 1260  
QY 1261 TACAGCAAGGCACTGAGGATTTCTTGTAGATCTGATGCTGTTGCTCGGCTTGAATCC 1320  
DB 1261 TACAGCAAGGCACTGAGGATTTCTTGTAGATCTGATGCTGTTGCTCGGCTTGAATCC 1320  
QY 1321 CCTGCCAAGACTAGCTCCATGGAAGAACTTCTCATCAAAAGCAAGAGCTTACAGAC 1380  
DB 1321 CCTGCCAAGACTAGCTCCATGGAAGAACTTCTCATCAAAAGCAAGAGCTTACAGAC 1380  
QY 1381 TCTCAGACAGATGCTCAAGATGAGCAAGAAATGACCCGTTACATCGAGATGCTCA 1440  
DB 1381 TCTCAGACAGATGCTCAAGATGAGCAAGAAATGACCCGTTACATCGAGATGCTCA 1440  
QY 1441 GAGGTGAGAGCTGTCTTATGATGAAGAGGCTGAGAGGCTTCAAGAGCTTCAAGAG 1500  
DB 1441 GAGGTGAGAGCTGTCTTATGATGAAGAGGCTGAGAGGCTTCAAGAGCTTCAAGAG 1500  
QY 1501 TCCCTTCTGAGACAGAGCTTGTCTACATCAACAGAAATGAGTAAAGGGAAT 1560  
DB 1501 TCCCTTCTGAGACAGAGCTTGTCTACATCAACAGAAATGAGTAAAGGGAAT 1560  
QY 1561 TTGAGAGCAAGCAAGATGAGGCTGCTCCAGAGAGTGAACAAGCACTGAGCTTCCAT 1620  
DB 1561 TTGAGAGCAAGCAAGATGAGGCTGCTCCAGAGAGTGAACAAGCACTGAGCTTCCAT 1620  
QY 1621 GATATCAGAGACAGAGCTGAGAGCTCCAGAAATCAAGAGAGATCCAGAGCTCA 1680  
DB 1621 GATATCAGAGACAGAGCTGAGAGCTCCAGAAATCAAGAGAGATCCAGAGCTCA 1680  
QY 1681 GTGAGAAAGATGAGGTTGATGATGATGATGATGATGATGATGATGATGATG 1740  
DB 1681 GTGAGAAAGATGAGGTTGATGATGATGATGATGATGATGATGATGATGATG 1740  
QY 1741 CCGAGTGAATCTTACGAATCTGAGCTGAGAGAGTCTCGGCTTGTGCTGAAGATTCAAG 1800  
DB 1741 CCGAGTGAATCTTACGAATCTGAGCTGAGAGAGTCTCGGCTTGTGCTGAAGATTCAAG 1800  
QY 1801 CCGAAAGCGACAGAAATGTCAGATTAACCTTTGAAGGCTTAAGATCAAGAGGCTTGA 1860  
DB 1801 CCGAAAGCGACAGAAATGTCAGATTAACCTTTGAAGGCTTAAGATCAAGAGGCTTGA 1860  
QY 1861 GTGGAGAAATATGAGAACTGAGAGAAATGATGCTGAGAGAGCTCAAAATTCAGAG 1920  
DB 1861 GTGGAGAAATATGAGAACTGAGAGAAATGATGCTGAGAGAGCTCAAAATTCAGAG 1920  
QY 1921 CTCGAAGAGAACTGAGAGAGCTTAAAGGCTGAGAGAGCTGAGAGAGCTGCTGAG 1980  
DB 1921 CTCGAAGAGAACTGAGAGAGCTTAAAGGCTGAGAGAGCTGAGAGAGCTGCTGAG 1980  
QY 1981 AATATCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
DB 1981 AATATCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
QY 2041 GATTCCTTGAAGGATCAAGAAAGAGCTGAGAGAGCTGAGAGAGAGCTGAGAGAG 2100  
DB 2041 GATTCCTTGAAGGATCAAGAAAGAGCTGAGAGAGCTGAGAGAGAGCTGAGAGAG 2100  
QY 2101 GAGAACAGGTTAAGATGAGACATGAGAGCTTGAAGAAACAGACTGAAGATGAC 2160

Db 2101 GAGAACAAAGGTAAGAGACTAGAGACCATGAGCCGTAGAGAAAACAGACTGAGAGTAC 2160  
Qy 2161 ATCCGACAAAATCCCAACAGATCCAGCAGATGGCTGATAAATTTCTGAGCTCGAAGAG 2220  
Db 2161 ATCCGACAAAATCCCAACAGATCCAGCAGATGGCTGATAAATTTCTGAGCTCGAAGAG 2220  
Qy 2221 AAACATCCGGAGAGCCCAAGTCTCAGCCCGACCTTGAAGTGCACCTGAAACAGAAAGAG 2280  
Db 2221 AAACATCCGGAGAGCCCAAGTCTCAGCCCGACCTTGAAGTGCACCTGAAACAGAAAGAG 2280  
Qy 2281 CAGCACTATGAGAGAAAAGATTAAAGTGTGACCAATCGATAGATAAGAAAAGACTGGCTAG 2340  
Db 2281 CAGCACTATGAGAGAAAAGATTAAAGTGTGACCAATCGATAGATAAGAAAAGACTGGCTAG 2340  
Qy 2341 AAGAGACACTGAGAGAACATGATGAGACACAGAGAGAGAGCCCATGAGAGAGGCAAA 2400  
Db 2341 AAGAGACACTGAGAGAACATGATGAGACACAGAGAGAGAGCCCATGAGAGAGGCAAA 2400  
Qy 2401 ATTCTCAGCCGACAGAGAGCCGATGATCAATGCTATGATGATCCAAATCGATCCCTGAA 2460  
Db 2401 ATTCTCAGCCGACAGAGAGCCGATGATCAATGCTATGATGATCCAAATCGATCCCTGAA 2460  
Qy 2461 CAGAGGATTTGAGAACTGCTGAAAGCCAAATAACTTGCAGCAATAGCAGTCTTTTAAAC 2520  
Db 2461 CAGAGGATTTGAGAACTGCTGAAAGCCAAATAACTTGCAGCAATAGCAGTCTTTTAAAC 2520  
Qy 2521 CAAAGGAACTAGAGAGCCCAAGAGAGATGATTTCTGAACTCGAGCAA CAGAAATTTTAC 2580  
Db 2521 CAAAGGAACTAGAGAGCCCAAGAGAGATGATTTCTGAACTCGAGCAA CAGAAATTTTAC 2580  
Qy 2581 CTGGAGACACAGAGGCTGGAGAAAGTTGAGAGCCCAAGAACTGGAAGAGAGAGAGTGGAG 2640  
Db 2581 CTGGAGACACAGAGGCTGGAGAAAGTTGAGAGCCCAAGAACTGGAAGAGAGAGAGTGGAG 2640  
Qy 2641 AAGATCAGCCACCAAGACCAACAGTGAACAAGATCGGCTGCTGAACTGGAACAAGATTG 2700  
Db 2641 AAGATCAGCCACCAAGACCAACAGTGAACAAGATCGGCTGCTGAACTGGAACAAGATTG 2700  
Qy 2701 CCGGAGGTCAGTCTAGAGCAGAGAGAGAGAACTGAGCTCAAGCCGCAAGCTCAGAGAG 2760  
Db 2701 CCGGAGGTCAGTCTAGAGCAGAGAGAGAGAACTGAGCTCAAGCCGCAAGCTCAGAGAG 2760  
Qy 2761 CTACAGCTCTCCCTGAGAGAGGCGAGTCAAGTTGAGAGCCCTGCAAGGCTGCAAGGCG 2820  
Db 2761 CTACAGCTCTCCCTGAGAGAGGCGAGTCAAGTTGAGAGCCCTGCAAGGCTGCAAGGCG 2820  
Qy 2821 GCCCTGAGAGCCAGCTTCCGAGGCGAAGACAGAGCTGGAAGAGACCAAGCAGAGAGCT 2880  
Db 2821 GCCCTGAGAGCCAGCTTCCGAGGCGAAGACAGAGCTGGAAGAGAGACCAAGCAGAGAGCT 2880  
Qy 2881 GAAAGAGAGATCCAGGACCTCAGGCACTATGAGTGAATCCAGCCGCAAAATTTGATGCT 2940  
Db 2881 GAAAGAGAGATCCAGGACCTCAGGCACTATGAGTGAATCCAGCCGCAAAATTTGATGCT 2940  
Qy 2941 CTTGTGAACAGCTGTACTGTATGACAGACTGAGAGGAGCATTAAACCACTAGAGAGCT 3000  
Db 2941 CTTGTGAACAGCTGTACTGTATGACAGACTGAGAGGAGCATTAAACCACTAGAGAGCT 3000  
Qy 3001 GACAAAGCTGAACTCAACCAAACTTCTACTTGTCCAAACCACTGATGAGAGCTTCT 3060  
Db 3001 GACAAAGCTGAACTCAACCAAACTTCTACTTGTCCAAACCACTGATGAGAGCTTCT 3060  
Qy 3061 GGGCCCAACGACGAGATTGTACAACTCGAAGTGAAGCCATCTCCGCGGGAATC 3120  
Db 3061 GGGCCCAACGACGAGATTGTACAACTCGAAGTGAAGCCATCTCCGCGGGAATC 3120  
Qy 3121 ACGGAGCAGAGATGACAGCTTACAGCCAGAGCAAAAGATGAGAGCTGGAAGACAG 3180  
Db 3121 ACGGAGCAGAGATGACAGCTTACAGCCAGAGCAAAAGATGAGAGCTGGAAGACAG 3180  
Qy 3181 TGCAACCATGCTGAGAGAACAGTCTATGATTTGAGAGCCCTTAAACGATGAGCTGCTGAA 3240  
Db 3181 TGCAACCATGCTGAGAGAACAGTCTATGATTTGAGAGCCCTTAAACGATGAGCTGCTGAA 3240

Qy 3241 AAAGACCGCAGTGGAGAGGCTTGAGAGAGCGTCTGGGTATGATGAATAATCCAGTTTGA 3300  
Db 3241 AAAGACCGCAGTGGAGAGGCTTGAGAGAGCGTCTGGGTATGATGAATAATCCAGTTTGA 3300  
Qy 3301 TGTGGGTTTGAAGCTGCAAGAGATGCTGACACCCGAGAAACAGAGCAGAGGCGAGAGCC 3360  
Db 3301 TGTGGGTTTGAAGCTGCAAGAGATGCTGACACCCGAGAAACAGAGCAGAGGCGAGAGCC 3360  
Qy 3361 GATCAGCCGATACCCGAGTCTTCCAGAGTGTGAGCTGGAGCTGGAAGAGACCAAGGCT 3420  
Db 3361 GATCAGCCGATACCCGAGTCTTCCAGAGTGTGAGCTGGAGCTGGAAGAGACCAAGGCT 3420  
Qy 3421 GAGATTCTCGCTCTGAGAGGCTCTCAAGAGCAGAGCTGAAGCCGAGAGCCCTCTCT 3480  
Db 3421 GAGATTCTCGCTCTGAGAGGCTCTCAAGAGCAGAGCTGAAGCCGAGAGCCCTCTCT 3480  
Qy 3481 GACAAAGCTCAATGACCTGAGAGAGATGATGCTTGAATGAATGCCCCGAAGCTTAA 3540  
Db 3481 GACAAAGCTCAATGACCTGAGAGAGATGATGCTTGAATGAATGCCCCGAAGCTTAA 3540  
Qy 3541 CAGCAGAGCTGAGAGCTGAAAGAGAGCTCAAAACAGAGGCTTCTGAAAGACCAAGC 3600  
Db 3541 CAGCAGAGCTGAGAGCTGAAAGAGAGCTCAAAACAGAGGCTTCTGAAAGACCAAGC 3600  
Qy 3601 TTACAGCAGCAGATGAGCCTGCAAGAAATCAACTTTCCGCTGACTCAAGACCTGCA 3660  
Db 3601 TTACAGCAGCAGATGAGCCTGCAAGAAATCAACTTTCCGCTGACTCAAGACCTGCA 3660  
Qy 3661 GAAAGCTTGAATCGGAGCTGATCTAAGAGACAGAAAGAGTGAAGCTTGGAGTATGAGCTG 3720  
Db 3661 GAAAGCTTGAATCGGAGCTGATCTAAGAGACAGAAAGAGTGAAGCTTGGAGTATGAGCTG 3720  
Qy 3721 GAAAACATTCAGGTTCTCTATTTCTATGAAAGAGTGAAGAAAGAGCACTATTTCTCA 3780  
Db 3721 GAAAACATTCAGGTTCTCTATTTCTATGAAAGAGTGAAGAAAGAGCACTATTTCTCA 3780  
Qy 3781 CAAACCAAACTCATGATTTTCTGAAAGCCAAATGAGCAACCTGCTAAAAAGAAAAG 3840  
Db 3781 CAAACCAAACTCATGATTTTCTGAAAGCCAAATGAGCAACCTGCTAAAAAGAAAAG 3840  
Qy 3841 GTTCCCTGCAAGTACATGAGCTGAGAGCTGAGGCTGAGAGAGAGAGAGTCTGCTGCA 3900  
Db 3841 GTTCCCTGCAAGTACATGAGCTGAGAGCTGAGGCTGAGAGAGAGAGAGTCTGCTGCA 3900  
Qy 3901 GAGCTAGAGAGAGGCTTCCAGAAAGCCCGCATGAGCTCCGCTCCGCGGAGAGAGAGCT 3960  
Db 3901 GAGCTAGAGAGAGGCTTCCAGAAAGCCCGCATGAGCTCCGCTCCGCGGAGAGAGAGCT 3960  
Qy 3961 GCCCAGCGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020  
Db 3961 GCCCAGCGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020  
Qy 4021 ATGCGCATGTCGCGCATGTCGAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080  
Db 4021 ATGCGCATGTCGCGCATGTCGAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080  
Qy 4081 GCGCCGCGCATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140  
Db 4081 GCGCCGCGCATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140  
Qy 4141 AAGGAGCAGATGAGCAGCAATATTTCTCAAGGATTCAGAGTGAAGTGAAGTGAAGG 4200  
Db 4141 AAGGAGCAGATGAGCAGCAATATTTCTCAAGGATTCAGAGTGAAGTGAAGTGAAGG 4200  
Qy 4201 ACAAAGTGTGCTGTGTGTCTGAGATCCGTGACCTTTGAGAGAGAGAGAGAGAGAG 4260  
Db 4201 ACAAAGTGTGCTGTGTGTCTGAGATCCGTGACCTTTGAGAGAGAGAGAGAGAGAG 4260  
Qy 4261 GAAATGTAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4320  
Db 4261 GAAATGTAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4320

QY 4321 CCTGCTGATATGACACACTTCAACGAGGCTTCTCCGCGTGAACAAATGAACTCCCA 4380  
 Db 4321 CCTGCTGATATGACACACTTCAACGAGGCTTCTCCGCGTGAACAAATGAACTCCCA 4380  
 QY 4381 GGTCTCCAGACCAAGAGCCGAGAGAGCTTGAACCTGGAAGGGTGAAGAGTGC 4440  
 Db 4381 GGTCTCCAGACCAAGAGCCGAGAGAGCTTGAACCTGGAAGGGTGAAGAGTGC 4440  
 QY 4441 AGGAATTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4500  
 Db 4441 AGGAATTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4500  
 QY 4501 AAGTCTCAATTTATGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560  
 Db 4501 AAGTCTCAATTTATGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560  
 QY 4561 GAGCTGCTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620  
 Db 4561 GAGCTGCTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620  
 QY 4621 GCAATTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
 Db 4621 GCAATTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
 QY 4681 ACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4740  
 Db 4681 ACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4740  
 QY 4741 TGGGTCAACGCTTGAATGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800  
 Db 4741 TGGGTCAACGCTTGAATGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800  
 QY 4801 GGTGATGCTAACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
 Db 4801 GGTGATGCTAACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
 QY 4861 ATGAATGACAGCTGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920  
 Db 4861 ATGAATGACAGCTGCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920  
 QY 4921 TACGCTGATGATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4980  
 Db 4921 TACGCTGATGATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4980  
 QY 4981 CAAATTTATATTTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5040  
 Db 4981 CAAATTTATATTTATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5040  
 QY 5041 CTGTGTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5100  
 Db 5041 CTGTGTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5100  
 QY 5101 CAGCCGACATCTCAACCAATTTTGAAGCTGTCAAGGCTGCCACTTGTGTGGGCA 5160  
 Db 5101 CAGCCGACATCTCAACCAATTTTGAAGCTGTCAAGGCTGCCACTTGTGTGGGCA 5160  
 QY 5161 GGCAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5220  
 Db 5161 GGCAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5220  
 QY 5221 CGCTCAACGAAAGCTTCAAGAAATCTGATCCGAGAAAGATAGAGCTTCAGAGGCC 5280  
 Db 5221 CGCTCAACGAAAGCTTCAAGAAATCTGATCCGAGAAAGATAGAGCTTCAGAGGCC 5280  
 QY 5281 TGCAGCTGATCACTTCAACCAATTTAGAGTATCTTCAATGGAACCAATTAATTTCA 5340  
 Db 5281 TGCAGCTGATCACTTCAACCAATTTAGAGTATCTTCAATGGAACCAATTAATTTCA 5340  
 QY 5341 ATCGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5400  
 Db 5341 ATCGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5400  
 QY 5401 CCTGCTGTGTTTGGCGCTCTTCCAAAGCTTCCCTGTCTCAATGCTGAGGTGAACAGC 5460

Db 5401 CCTGCTGTGTTTGGCGCTCTTCCAAAGCTTCCCTGTCTCAATGCTGAGGTGAACAGC 5460  
 QY 5461 GCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5520  
 Db 5461 GCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5520  
 QY 5521 TACGAAAGAGTACCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5580  
 Db 5521 TACGAAAGAGTACCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5580  
 QY 5581 TACGAAAGAGTACCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5640  
 Db 5581 TACGAAAGAGTACCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5640  
 QY 5641 GCAGCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5700  
 Db 5641 GCAGCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5700  
 QY 5701 CTGGGCTCTGCAATTTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5760  
 Db 5701 CTGGGCTCTGCAATTTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5760  
 QY 5761 AGGCTCAATTTGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5820  
 Db 5761 AGGCTCAATTTGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5820  
 QY 5821 CCGTCCACTCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5880  
 Db 5821 CCGTCCACTCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5880  
 QY 5881 ACCAAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5940  
 Db 5881 ACCAAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5940  
 QY 5941 AGCAGACCTCCAGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6000  
 Db 5941 AGCAGACCTCCAGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6000  
 QY 6001 CGCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6060  
 Db 6001 CGCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6060  
 QY 6061 CCGGAGAGGCTGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6120  
 Db 6061 CCGGAGAGGCTGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6120  
 QY 6121 CCGCTGTCCAGGTGAACAAAGGTCTGGAGACAGTCTTCAATATA 6165  
 Db 6121 CCGCTGTCCAGGTGAACAAAGGTCTGGAGACAGTCTTCAATATA 6165  
 RESULT 4  
 AAL55215  
 ID AAL55215 standard; DNA; 8603 BP.  
 XX AAL55215;  
 AC 01-MAY-2003 (first entry)  
 XX  
 DE Human CRK related DNA sequence, SEQ ID No 4.  
 XX  
 XX Anorectic; hypotensive; cardiast; antilipemic; cerebroprotective;  
 KW antiagout; osteopathic; antidiabetic; cytoprotective; antidepressant;  
 KW immunomodulator; antineoplastic; tranquilizer; antiparkinsonian; nootropic;  
 KW neuroprotective; antineoplastic; antidiabetic; analgesic;  
 KW human cation rho/rac-interacting kinase; enzyme; CRK; ameliorating;  
 KW obesity; comorbidities; cancer; anorexia; cachexia; bulimia;  
 KW central nervous system disorder; chronic obstructive pulmonary disease;  
 KW diabetes; pain; de.  
 XX  
 XX Homo sapiens.  
 OS

PN MO2003004523-AL.  
XX  
PD 16-JAN-2003.  
XX  
PF 28-JUN-2002; 2002MO-EP07156.  
XX  
PR 02-JUL-2001; 2001US-301841P.  
PR 11-DEC-2001; 2001US-338651P.  
PR 25-APR-2002; 2002US-375014P.  
XX  
PA (FARB ) BAYER AG.  
XX  
PI Zhu Z;  
XX  
DR WPI; 2003-221576/21.  
XX  
PT New human citron rho/rac-interacting kinase (CRIK) polypeptide and  
PT polynucleotide, useful in preventing, ameliorating or treating diseases  
PT associated with human CRIK dysfunction, e.g. obesity, diabetes or  
PT Alzheimer's disease  
XX  
PS Disclosure, Fig 4; 237p; English.  
XX  
CC The invention relates to an isolated polynucleotide encoding a human  
CC citron rho/rac-interacting kinase polypeptide. The isolated  
CC polynucleotide comprises a 6165 or 8603 base pair sequence, given in the  
CC specification. The human citron rho/rac-interacting kinase (CRIK)  
CC polypeptide and polynucleotide are useful in preventing, ameliorating, or  
CC treating diseases associated with human CRIK dysfunction such as obesity  
CC and obesity-associated comorbidities (e.g. hypertension, coronary artery  
CC disease, hyperlipidemia, stroke, gout, osteoarthritis, some types of  
CC cancer including endometrial, breast, prostate and colon cancer),  
CC anorexia, cachexia, bulimia, central nervous system disorders (e.g. mood  
CC disorders), anxiety disorders, Parkinson's disease, or Alzheimer's  
CC disease), chronic obstructive pulmonary disease, or diabetes. These can  
CC also be used to treat pain associated with the disorders. The human CRIK  
CC polypeptide is also useful in diagnostic assays or in genetic testing.  
CC The expression vector or the reagent is useful in preparing a medicament  
CC for modulating the activity of a human CRIK in a disease, e.g. obesity, a  
CC central nervous system disorder, or chronic obstructive pulmonary  
CC disease. The fusion protein is useful for generating antibodies against a  
CC CRIK polypeptide and for use in various assay systems. The methods are  
CC useful in producing and detecting the polynucleotide and polypeptide and  
CC in screening for agents that modulate the activity of the human CRIK  
CC polypeptide. This polynucleotide sequence represents a DNA sequence  
CC relating to the human CRIK protein of the invention.  
XX  
XX Sequence 8603 BP; 2305 A; 2206 C; 2215 G; 1877 T; 0 other;  
SQ  
Query Match 99.9%; Score 6160.2; DB 25; Length 8603;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6162; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAGTTGAAGTCAATATATGAGCGCGAATCCTTTGGATGCTGTCGTGAACCCATT 60  
DB 1 AAGTTGAAGTCAATATATGAGCGCGAATCCTTTGGATGCTGTCGTGAACCCATT 60  
QY 61 GCCAGCGCGGCGCTCCAGGCTGATCTGTTCTTCCAGGCGAAACACCCCTTATGACTCAA 120  
DB 61 GCCAGCGCGGCGCTCCAGGCTGATCTGTTCTTCCAGGCGAAACACCCCTTATGACTCAA 120  
QY 121 CAGCAGATGCTCTCTTTTCCGAGAAAGGATATTAGATGCCCTCTTGTCTCTTTGAA 180  
DB 121 CAGCAGATGCTCTCTTTTCCGAGAAAGGATATTAGATGCCCTCTTGTCTCTTTGAA 180  
QY 181 GAATGCACTGCTGCTCTGATGAAGATTAAAGCAGTGAACAATTGTCGGAAGTAT 240  
DB 181 GAATGCACTGCTGCTCTGATGAAGATTAAAGCAGTGAACAATTGTCGGAAGTAT 240  
QY 241 TCCGACACCATAGCTGAGTTACAGAGCTCCAGCCCTTCGGCAAGAACTTCCGAAGTCAA 300  
DB 241 TCCGACACCATAGCTGAGTTACAGAGCTCCAGCCCTTCGGCAAGAACTTCCGAAGTCAA 300

QY 301 AGCTCTGATAGCTGTGTGCTCACTTTGCTGAAGTCAGTGTGTTAAGAGAAAGCAACCGG 360  
DB 301 AGCTCTGATAGCTGTGTGCTCACTTTGCTGAAGTCAGTGTGTTAAGAGAAAGCAACCGG 360  
QY 361 GACATCATGCTATGAAGATGAAGAAAGAGCTTTATGGCCCGAGGAGCTTTCA 420  
DB 361 GACATCATGCTATGAAGATGAAGAAAGAGCTTTATGGCCCGAGGAGCTTTCA 420  
QY 421 TTTTGGAGAAAGCGGAAACATATTATCTGAGAGCAAGCCGTGATCCCAATTA 480  
DB 421 TTTTGGAGAAAGCGGAAACATATTATCTGAGAGCAAGCCGTGATCCCAATTA 480  
QY 481 CAGTATGCCCTTTAGAGCAAAAATCACCTTTATCTGCTCATGGAATATCAGCTGAGGG 540  
DB 481 CAGTATGCCCTTTAGAGCAAAAATCACCTTTATCTGCTCATGGAATATCAGCTGAGGG 540  
QY 541 GACTTGCTGCACTTTTGAATATATAGAGACAGTATATGAAACCTGATCACTTT 600  
DB 541 GACTTGCTGCACTTTTGAATATATAGAGACAGTATATGAAACCTGATCACTTT 600  
QY 601 TACCTAGCTGAGCTGATTTTGGCTGTCAAGGCTTCATCTGATGGATACGTGATCGA 660  
DB 601 TACCTAGCTGAGCTGATTTTGGCTGTCAAGGCTTCATCTGATGGATACGTGATCGA 660  
QY 661 GACATCAAGCTGAGAAACATTTCTGTTGACCGCAGAGACATCAAGCTGTGATTTT 720  
DB 661 GACATCAAGCTGAGAAACATTTCTGTTGACCGCAGAGACATCAAGCTGTGATTTT 720  
QY 721 GGAATCGCGCGGAAATGAAATTCAAAACAAGATGTAATGCCAACTCCGATTTGGAGC 780  
DB 721 GGAATCGCGCGGAAATGAAATTCAAAACAAGATGTAATGCCAACTCCGATTTGGAGC 780  
QY 781 CCAGATTATAGCTCTCTGAGAGTCTGACTGTGATGAACGGAGATGAAAGACACTAC 840  
DB 781 CCAGATTATAGCTCTCTGAGAGTCTGACTGTGATGAACGGAGATGAAAGACACTAC 840  
QY 841 GGCCTGAGCTGAGTCTGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 900  
DB 841 GGCCTGAGCTGAGTCTGCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAG 900  
QY 901 TCCGCTTCCGAGAGGAAACCTGCGCAGAACCTTCAATTAACCTTATGAAATTTCCAGCG 960  
DB 901 TCCGCTTCCGAGAGGAAACCTGCGCAGAACCTTCAATTAACCTTATGAAATTTCCAGCG 960  
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020  
DB 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020  
QY 1021 TTGTTGTGGGCGCAAAAGAGACTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1080  
DB 1021 TTGTTGTGGGCGCAAAAGAGACTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1080  
QY 1081 TCTTAAATGACTGGAACAACATTTGTAATCTCTCCCTCCCTGCTTCCCACTGTAAG 1140  
DB 1081 TCTTAAATGACTGGAACAACATTTGTAATCTCTCCCTCCCTGCTTCCCACTGTAAG 1140  
QY 1141 TCTGAGATGACACTTCAATTTTGAATGAACAGAAAGATTTGAGGTTTTCATCTCT 1200  
DB 1141 TCTGAGATGACACTTCAATTTTGAATGAACAGAAAGATTTGAGGTTTTCATCTCT 1200  
QY 1201 CCGTGCAGCTGAGCCCTCAAGCTTCTGGGTGAAGAACTGCGGTTTGGGGTTTTCG 1260  
DB 1201 CCGTGCAGCTGAGCCCTCAAGCTTCTGGGTGAAGAACTGCGGTTTGGGGTTTTCG 1260  
QY 1261 TACAGCAAGGCACTGGGAGATTCTTGGTATGATCTGATCTGATCTGATCTGATCTG 1320  
DB 1261 TACAGCAAGGCACTGGGAGATTCTTGGTATGATCTGATCTGATCTGATCTGATCTG 1320  
QY 1321 CCTGCCAAGACTAGCTTCATGAGAAAGAACTTCTCATCAAAAGCAAGAGCTTACAGAG 1380  
DB 1321 CCTGCCAAGACTAGCTTCATGAGAAAGAACTTCTCATCAAAAGCAAGAGCTTACAGAG 1380  
QY 1381 TCTCAGAGCAAGTGTCACAAGATGAGAGCAAGAAATGACCCGGTTACATGAGAGTGTCA 1440



```
Db 1381 TCTCAGGACAAATGTCACAAAGATGGAGCGAGAAATGACCCGGTTACATCGAGAGTGTCA 1440
Qy 1441 GAGGTGGAGGCTGTGCTTATGTCAGAAAGAGGTGAGCTGAAGCCCTCTGAGACTCAAGA 1500
Db 1441 GAGGTGGAGGCTGTGCTTATGTCAGAAAGAGGTGAGCTGAAGCCCTCTGAGACTCAAGA 1500
Qy 1501 TCCCTCTGGAGCAGAGACCTTGCTACCTACATCAGAAATGCAATGATAGCTTAAAGGAACT 1560
Db 1501 TCCCTCTGGAGCAGAGACCTTGCTACCTACATCAGAAATGCAATGATAGCTTAAAGGAACT 1560
Qy 1561 TTGGAGCAAGACCGATGGAGGTGTCCAGAGAGATGACAAAGCACTGACGCTTCCAT 1620
Db 1561 TTGGAGCAAGACCGATGGAGGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCAT 1620
Qy 1621 GATATCAGAGCAGAGACCGGAAAGCTCCAGAAATCAAGAGCAGAGAGTCCAGGCTCA 1680
Db 1621 GATATCAGAGCAGAGACCGGAAAGCTCCAGAAATCAAGAGCAGAGAGTCCAGGCTCA 1680
Qy 1681 GTGGAGAAATGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
Db 1681 GTGGAGAAATGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
Qy 1741 CGGAGTGATCTCTACGAATCTGAGCTGAGAGGTCTCGGCTTGCTGCTGAGAAATTCAG 1800
Db 1741 CGGAGTGATCTCTACGAATCTGAGCTGAGAGGTCTCGGCTTGCTGCTGAGAAATTCAG 1800
Qy 1801 CGGAAAGCAGAGATGTCACATTAATCTGTTGAAGCTTAAGATCAAGGAAAGCTTGA 1860
Db 1801 CGGAAAGCAGAGATGTCACATTAATCTGTTGAAGCTTAAGATCAAGGAAAGCTTGA 1860
Qy 1861 GTGGAGAAATGATGGAAATCTGAGAAAGTCAATGTGAGCAGAGCTTCAAAATTCAGAG 1920
Db 1861 GTGGAGAAATGATGGAAATCTGAGAAAGTCAATGTGAGCAGAGCTTCAAAATTCAGAG 1920
Qy 1921 CTCCAGAGAAATCTGAGAAAGCTGTAAAGCCAGCACTGAGGCACTGAGCTGCTGAG 1980
Db 1921 CTCCAGAGAAATCTGAGAAAGCTGTAAAGCCAGCACTGAGGCACTGAGCTGCTGAG 1980
Qy 1981 AATATCCGCGCAAGGCAAGGACCGAGCGAGGAGAGCTGAGAAAGCTGCAAAACCGAG 2040
Db 1981 AATATCCGCGCAAGGCAAGGACCGAGCGAGGAGAGCTGAGAAAGCTGCAAAACCGAG 2040
Qy 2041 GATTCCTCTGAAAGCATCAAAAAGAGCTGGTGAAGCTGAGAAACCCCGCATTTCTG 2100
Db 2041 GATTCCTCTGAAAGCATCAAAAAGAGCTGGTGAAGCTGAGAAACCCCGCATTTCTG 2100
Qy 2101 GAGAACAGGTAAAGAGACTAGAGCCATGAGCGGTAGAGAAACAGACTGAGAGTAC 2160
Db 2101 GAGAACAGGTAAAGAGACTAGAGCCATGAGCGGTAGAGAAACAGACTGAGAGTAC 2160
Qy 2161 ATCCAGCAAAATCCCAACAGATCCAGCAAGTGGCTGATAAATTCGAGCTGAGAG 2220
Db 2161 ATCCAGCAAAATCCCAACAGATCCAGCAAGTGGCTGATAAATTCGAGCTGAGAG 2220
Qy 2221 AAACATCGGAGGCCCAAGTCTAGCCAGCACTTAAGTGAAGCTGAAACGAAAGAG 2280
Db 2221 AAACATCGGAGGCCCAAGTCTAGCCAGCACTTAAGTGAAGCTGAAACGAAAGAG 2280
Qy 2281 CAGCACTATGAGAAAAGATTAAAGTGTGAGCAATCAGATTAAGAAAGACTGCTGAC 2340
Db 2281 CAGCACTATGAGAAAAGATTAAAGTGTGAGCAATCAGATTAAGAAAGACTGCTGAC 2340
Qy 2341 AAGGAGCACTGAGAAACATGATGAGAGCAGAGAGAGAGGCCATGAGAAAGGCAAA 2400
Db 2341 AAGGAGCACTGAGAAACATGATGAGAGCAGAGAGAGAGGCCATGAGAAAGGCAAA 2400
Qy 2401 ATTCTCAGGCAACAGAGGAGATCAATGCTATGATTCAGAGATCGATCCCTGGA 2460
Db 2401 ATTCTCAGGCAACAGAGGAGATCAATGCTATGATTCAGAGATCGATCCCTGGA 2460
Qy 2461 CAGAGATTGTGAACTGTCTGAAGCAATAAATTGACAGAAATGACAGTCTTTTACC 2520
Db 2461 CAGAGATTGTGAACTGTCTGAAGCAATAAATTGACAGAAATGACAGTCTTTTACC 2520
Db 2461 CAGAGATTGTGAACTGTCTGAAGCAATAAATTGACAGAAATGACAGTCTTTTACC 2520
Qy 2521 CAAAGAAACATGAAAGGCCCAAGAGAGATGATTTCTGAATCAGGCAACAGAAATTTAC 2580
Db 2521 CAAAGAAACATGAAAGGCCCAAGAGAGATGATTTCTGAATCAGGCAACAGAAATTTAC 2580
Qy 2581 CTGAGACACAGGCTGGGAAATTGAGAGCCAGAAACCGAAACCTGAGAGAGAGCTGAG 2640
Db 2581 CTGAGACACAGGCTGGGAAATTGAGAGCCAGAAACCGAAACCTGAGAGAGAGCTGAG 2640
Qy 2641 AAGATCAGCCACCAAGACACAGATGACAAAGATCCGCTGCTGAACTGAGACCAAGATTG 2700
Db 2641 AAGATCAGCCACCAAGACACAGATGACAAAGATCCGCTGCTGAACTGAGACCAAGATTG 2700
Qy 2701 CGGAGGTCAAGCTTGAAGACAGAGAGAGAAATCTGAGCTTAAAGGCTCAACAGAG 2760
Db 2701 CGGAGGTCAAGCTTGAAGACAGAGAGAGAAATCTGAGCTTAAAGGCTCAACAGAG 2760
Qy 2761 CTACAGCTCTCCCTGCAAGAGCGAGTCAAGTTGACAGCCCTGCAAGGCTCAAGGCT 2820
Db 2761 CTACAGCTCTCCCTGCAAGAGCGAGTCAAGTTGACAGCCCTGCAAGGCTCAAGGCT 2820
Qy 2821 GCCCTGAGAGCCAGCTTCCGCAAGCCAGACAGAGCTGGAAGAGACCAAGAGAGCT 2880
Db 2821 GCCCTGAGAGCCAGCTTCCGCAAGCCAGACAGAGCTGGAAGAGACCAAGAGAGCT 2880
Qy 2881 GAAAGAGATCCAGGCACTCAGGCAATGAGATGAAATCCAGCGCAATTTGATGT 2940
Db 2881 GAAAGAGATCCAGGCACTCAGGCAATGAGATGAAATCCAGCGCAATTTGATGT 2940
Qy 2941 CTTGTAAACAGCTGATCTGTAAATCAAGACCTGAGAGAGAGCTTAAACAGCTGAG 3000
Db 2941 CTTGTAAACAGCTGATCTGTAAATCAAGACCTGAGAGAGAGCTTAAACAGCTGAG 3000
Qy 3001 GACAAAGCTGAACTCAACCAAACTTTTACTTGTCCAAACAACTGATGAGGCTTCT 3060
Db 3001 GACAAAGCTGAACTCAACCAAACTTTTACTTGTCCAAACAACTGATGAGGCTTCT 3060
Qy 3061 GCGCCCAACGAGATTTGACAACTGGAGAGTGAAGTGAACATCTCCGCGGAGATC 3120
Db 3061 GCGCCCAACGAGATTTGACAACTGGAGAGTGAAGTGAACATCTCCGCGGAGATC 3120
Qy 3121 ACGGAAAGAGATGAGCTTACACGACAGAGCAAAAGATGAGAGCTTGAAGACAG 3180
Db 3121 ACGGAAAGAGATGAGCTTACACGACAGAGCAAAAGATGAGAGCTTGAAGACAG 3180
Qy 3181 TGACCATGCTGAGAGAAACAGCTATGATTTGAGAGCCCTTAAACGATGAGCTGTA 3240
Db 3181 TGACCATGCTGAGAGAAACAGCTATGATTTGAGAGCCCTTAAACGATGAGCTGTA 3240
Qy 3241 AAAGAGCGGAGTGGAGAGGCTGGAGAGGCTCTGGGTGATGAGAAATCCCAATTTGAG 3300
Db 3241 AAAGAGCGGAGTGGAGAGGCTGGAGAGGCTCTGGGTGATGAGAAATCCCAATTTGAG 3300
Qy 3301 TGTGGGTTGAGAGCTGAGAGATGCTGAGACACCGAGAAACAGAGAGAGGCGAGAGCC 3360
Db 3301 TGTGGGTTGAGAGCTGAGAGATGCTGAGACACCGAGAAACAGAGAGAGGCGAGAGCC 3360
Qy 3361 GATCAGCGATTCACCGAGTTCGCCAGGTGTGAGAGCTGGCATGAGAGAGCAAGGCT 3420
Db 3361 GATCAGCGATTCACCGAGTTCGCCAGGTGTGAGAGCTGGCATGAGAGAGCAAGGCT 3420
Qy 3421 GAGATTCTGCTGAGCAGGCTCTCAAGAGACAGAGCTGAAAGGCGAGAGCTTCT 3480
Db 3421 GAGATTCTGCTGAGCAGGCTCTCAAGAGACAGAGCTGAAAGGCGAGAGCTTCT 3480
Qy 3481 GACAACTCAATGACTTGAAGAAAGCATGCTATGCTTGAATGATGCTCCGAGGCTTA 3540
Db 3481 GACAACTCAATGACTTGAAGAAAGCATGCTATGCTTGAATGATGCTCCGAGGCTTA 3540
Qy 3541 CAGCAGAGCTGAGAGCTGAACAGAGCTCAACAGAGGCTTCTGGAAGAGCAAGCAAA 3600
Db 3541 CAGCAGAGCTGAGAGCTGAACAGAGCTCAACAGAGGCTTCTGGAAGAGCAAGCAAA 3600
```



QY 3601 TTACAGCAGAGATGAGCTGCGAGAAAAATCAATTTTCGTGACTGACAGACTGCAA 3660  
DB 3601 TTACAGCAGAGATGAGCTGCGAGAAAAATCAATTTTCGTGACTGACAGACTGCAA 3660  
QY 3661 GAAGCTCTAATGCGGCTGATTTACTAAGACAGAAAGAAAGTGAATCTTGGAGTACAGCTG 3720  
DB 3661 GAAGCTCTAATGCGGCTGATTTACTAAGACAGAAAGAAAGTGAATCTTGGAGTACAGCTG 3720  
QY 3721 GAAACATTCAGGTTCTCTATTCTCATGAAAAAGTGAAGTGAAGGCACTATTTCTCAA 3780  
DB 3721 GAAACATTCAGGTTCTCTATTCTCATGAAAAAGTGAAGTGAAGGCACTATTTCTCAA 3780  
QY 3781 CAAACCAAACTCATTTGATTTTCTGCAAGCCAAAATGAGCCACCTGTAAAAAGAAAAAG 3840  
DB 3781 CAAACCAAACTCATTTGATTTTCTGCAAGCCAAAATGAGCCACCTGTAAAAAGAAAAAG 3840  
QY 3841 GTTCCCTGAGATGACATGAGCTGAAGCTGGCCCTTGAGAGAGAAAGTCCCTGTGCA 3900  
DB 3841 GTTCCCTGAGATGACATGAGCTGAAGCTGGCCCTTGAGAGAGAAAGTCCCTGTGCA 3900  
QY 3901 GAGCTGAGAGAACCCCTTTCAGAAAGCCGATCGAGTCCGCTCCGCGGAGAGAGCT 3960  
DB 3901 GAGCTGAGAGAACCCCTTTCAGAAAGCCGATCGAGTCCGCTCCGCGGAGAGAGCT 3960  
QY 3961 GCCCACCAGAAAGCAACGACCAACCCATCCACGCAAGCCACCCGAGAGAGAG 4020  
DB 3961 GCCCACCAGAAAGCAACGACCAACCCATCCACGCAAGCCACCCGAGAGAGAG 4020  
QY 4021 ATGCGCATGTCGCGCATCGTGGGGTGGCGCAGAGACACAGCCGATCGATGAGCTGCTG 4080  
DB 4021 ATGCGCATGTCGCGCATCGTGGGGTGGCGCAGAGACACAGCCGATCGATGAGCTGCTG 4080  
QY 4081 GCCCGCCCATCCAGCCGAGAAAGAGTCTTCAACTCCAGAGAAATTTAGTCGCGCTTT 4140  
DB 4081 GCCCGCCCATCCAGCCGAGAAAGAGTCTTCAACTCCAGAGAAATTTAGTCGCGCTTT 4140  
QY 4141 AAGGAAGCATGACACACACATTTCTTCAACCGATTCACGATGAGACTGAAATGAGAGCC 4200  
DB 4141 AAGGAAGCATGACACACACATTTCTTCAACCGATTCACGATGAGACTGAAATGAGAGCC 4200  
QY 4201 ACAAAGTGTGCTGTGTGTGCGATACCGGCACTTTGAGCGCGAGCAATCCAAATGTC 4260  
DB 4201 ACAAAGTGTGCTGTGTGTGCGATACCGGCACTTTGAGCGCGAGCAATCCAAATGTC 4260  
QY 4261 GAAATGTCAGTGTGTGTGACCCCAAGTGTCTCCACGTCGTCGACGCACTTGGCGCTTG 4320  
DB 4261 GAAATGTCAGTGTGTGTGACCCCAAGTGTCTCCACGTCGTCGACGCACTTGGCGCTTG 4320  
QY 4321 CCTGCTGAATATGCCCAACATTTCAACCGAGGCTTTCTGCGGTGACAAATGAACTCCCA 4380  
DB 4321 CCTGCTGAATATGCCCAACATTTCAACCGAGGCTTTCTGCGGTGACAAATGAACTCCCA 4380  
QY 4381 GGTCTCCAGACCAAGGAGCCAGCAGCAGCTTGCACTTGAAAGGGTGAATGAAGTGCC 4440  
DB 4381 GGTCTCCAGACCAAGGAGCCAGCAGCAGCTTGCACTTGAAAGGGTGAATGAAGTGCC 4440  
QY 4441 AAGGAATTAACAAACGAGAGACAGAGCTGGGACAGGAATGATGCTCTGAGAGGATCA 4500  
DB 4441 AAGGAATTAACAAACGAGAGACAGAGCTGGGACAGGAATGATGCTCTGAGAGGATCA 4500  
QY 4501 AAAAGTCCCATTTAAGAAATGAAGCCAGAGAGCTGAGCAGAGCGCGGTGAGAAATTT 4560  
DB 4501 AAAAGTCCCATTTAAGAAATGAAGCCAGAGAGCTGAGCAGAGCGCGGTGAGAAATTT 4560  
QY 4561 GAGCTGTGCTTCCGAGCGGGATGATATTTCAATGAGTGGCTTGGTGTCTTCCGAATCTC 4620  
DB 4561 GAGCTGTGCTTCCGAGCGGGATGATATTTCAATGAGTGGCTTGGTGTCTTCCGAATCTC 4620  
QY 4621 GCAATTAACAGCCAAAGCAGATGTCCCATATCAATGAGATGAAATCTCAACCCGACACCC 4680  
DB 4621 GCAATTAACAGCCAAAGCAGATGTCCCATATCAATGAGATGAAATCTCAACCCGACACCC 4680

QY 4681 ACCGTGTGCGCCGGAGAAACCTCTATCTTGCTAGCTCCAGCTTCCCTGACAAACAGCGC 4740  
DB 4681 ACCGTGTGCGCCGGAGAAACCTCTATCTTGCTAGCTCCAGCTTCCCTGACAAACAGCGC 4740  
QY 4741 TGGGTACCGCCTTGAATATCATGTTGCGAGGTGGAGAGATTTCTAGGAAAAAGCAGAA 4800  
DB 4741 TGGGTACCGCCTTGAATATCATGTTGCGAGGTGGAGAGATTTCTAGGAAAAAGCAGAA 4800  
QY 4801 GCTGATGCTTAACTGCTTGGAAATCCCTGCTGTAACCTGGAAGGTATGACCGTCTAGAC 4860  
DB 4801 GCTGATGCTTAACTGCTTGGAAATCCCTGCTGTAACCTGGAAGGTATGACCGTCTAGAC 4860  
QY 4861 ATGAATGCAAGCTGCTTCAAGTGAACAGAGTGTGTTGGTGGGACCCGAGAGAGGCTC 4920  
DB 4861 ATGAATGCAAGCTGCTTCAAGTGAACAGAGTGTGTTGGTGGGACCCGAGAGAGGCTC 4920  
QY 4921 TACGCCCTGAATGTCTGAAAAATCTCCCTAACCAATGTCCTCCAGAAATTTGAGCAGTTC 4980  
DB 4921 TACGCCCTGAATGTCTGAAAAATCTCCCTAACCAATGTCCTCCAGAAATTTGAGCAGTTC 4980  
QY 4981 CAAATTTATATATCAAGGACCTGAGAGCTACTATGATAGCAGAGAGAGGCGGCA 5040  
DB 4981 CAAATTTATATATCAAGGACCTGAGAGCTACTATGATAGCAGAGAGAGGCGGCA 5040  
QY 5041 CTGTGTCTTGTGAGCGTGAAGAAAGTGAACAGTCCCTGGCCAGTCCCACTGCTGCC 5100  
DB 5041 CTGTGTCTTGTGAGCGTGAAGAAAGTGAACAGTCCCTGGCCAGTCCCACTGCTGCC 5100  
QY 5101 CAGCCCGACATCTCAACCCCAATTTTGAAGCTGTCAAGGGCTGCCACTTGTGGGGCA 5160  
DB 5101 CAGCCCGACATCTCAACCCCAATTTTGAAGCTGTCAAGGGCTGCCACTTGTGGGGCA 5160  
QY 5161 GGCAAGATTGAGACGGGCTGTGATCTGTGAGAGCATGCCAGCAAAAGTCGATCTC 5220  
DB 5161 GGCAAGATTGAGACGGGCTGTGATCTGTGAGAGCATGCCAGCAAAAGTCGATCTC 5220  
QY 5221 CGCTACAGAAAACTCTGACAAATCTGATCCGAGAAAGATAGAGACTTCAAGAGCC 5280  
DB 5221 CGCTACAGAAAACTCTGACAAATCTGATCCGAGAAAGATAGAGACTTCAAGAGCC 5280  
QY 5281 TGCAGCTGTATCCACTTCCAAATTAAGATTCATGATCTGATGGAACCAATTAATCTAGAA 5340  
DB 5281 TGCAGCTGTATCCACTTCCAAATTAAGATTCATGATCTGATGGAACCAATTAATCTAGAA 5340  
QY 5341 ATGCAATATGACAGATGACAGCTCGAGAGAAATCTGAGATGAAGATGCAATCTTGGCA 5400  
DB 5341 ATGCAATATGACAGATGACAGCTCGAGAGAAATCTGAGATGAAGATGCAATCTTGGCA 5400  
QY 5401 CCTGCTGTGTTGCGCCTCTTCCAAACAGCTTCCCTGTCTCAATCGTGCAGGTGAACAGC 5460  
DB 5401 CCTGCTGTGTTGCGCCTCTTCCAAACAGCTTCCCTGTCTCAATCGTGCAGGTGAACAGC 5460  
QY 5461 GAGAGGACCGAGAGAGATGACTTGTGTGTTCACAGAAATTTGAGAGTGTTCGTGATTTCT 5520  
DB 5461 GAGAGGACCGAGAGAGATGACTTGTGTGTTCACAGAAATTTGAGAGTGTTCGTGATTTCT 5520  
QY 5521 TACGGAAGAGCTGAGCGCAGACAGATCTCAAGTGAAGTGGCTTACCTTGGCTTGGC 5580  
DB 5521 TACGGAAGAGCTGAGCGCAGACAGATCTCAAGTGAAGTGGCTTACCTTGGCTTGGC 5580  
QY 5581 TACAGAGAAACCTATCTGTTGTGAACCACTTCAACTCACTGGAAGTATGAGATCCAG 5640  
DB 5581 TACAGAGAAACCTATCTGTTGTGAACCACTTCAACTCACTGGAAGTATGAGATCCAG 5640  
QY 5641 GAGCGCTCTGAGAGAGAGACCCCTGCGGAGGTACTGAGACATCCCGAACCCGCGCTAC 5700  
DB 5641 GAGCGCTCTGAGAGAGAGACCCCTGCGGAGGTACTGAGACATCCCGAACCCGCGCTAC 5700  
QY 5701 CTGGGCTCTGCAATTTCTCTAGAGAGATTTACTTGGCGTCTCTATCAACAGGATTAATTA 5760  
DB 5701 CTGGGCTCTGCAATTTCTCTAGAGAGATTTACTTGGCGTCTCTATCAACAGGATTAATTA 5760  
QY 5761 AAGGTCATTTGCTGCAAGGAAACCTCGTGAAGAGTCCGCGCACTGAACACCAACCGGGCC 5820

Db 5761 AGGTCATTCTGCTCAAGGAAACCTCGTGAAGAGTCCGGACATGAAACACACCGGGGC 5820  
 Qy 5821 CCGTCCACCTCCCGCAGAGGCCCAACAGAGAGGCCCAACCGATACAGAGACATC 5880  
 Db 5821 CCGTCCACCTCCCGCAGAGGCCCAACAGAGAGGCCCAACCGATACAGAGACATC 5880  
 Qy 5881 ACCAAGCGCGTGGCTCCAGGCCAGCGCCCGCCAGAGGCCCAACCGATACAGAGACATC 5940  
 Db 5881 ACCAAGCGCGTGGCTCCAGGCCAGCGCCCGCCAGAGGCCCAACCGATACAGAGACATC 5940  
 Qy 5941 AGCAGACCCCAACCGCTACCGCGAGGCGCGAGCCAGAGTGCAGAGGCAAAAGTCTCTGAGC 6000  
 Db 5941 AGCAGACCCCAACCGCTACCGCGAGGCGCGAGCCAGAGTGCAGAGGCAAAAGTCTCTGAGC 6000  
 Qy 6001 CGCCCTTGGAGAGAGAGAGTCCCGCGCGGATCTACAGACCGCGAGAGAGCGGTCC 6060  
 Db 6001 CGCCCTTGGAGAGAGAGTCCCGCGCGGATCTACAGACCGCGAGAGAGCGGTCC 6060  
 Qy 6061 CCCGCGAGGCTGTTTGAAGACAGAGAGAGGCGCGGCTGCGTGGGAGCGGTGAGACC 6120  
 Db 6061 CCCGCGAGGCTGTTTGAAGACAGAGAGAGGCGCGGCTGCGTGGGAGCGGTGAGACC 6120  
 Qy 6121 CGCGTGTCCAGAGTGAACAAAGTCTGGACCAAGTCTTCAAGTATA 6165  
 Db 6121 CGCGTGTCCAGAGTGAACAAAGTCTGGACCAAGTCTTCAAGTATA 6165

RESULT 5  
 AAL55217  
 ID AAL55217 standard; DNA; 6156 BP.  
 AC AAL55217;  
 XX  
 DT 01-MAY-2003 (first entry)  
 XX  
 DE Human CR1K related DNA sequence, SEQ ID No 8.  
 XX

XX Anorectic; hypotensive; cardiatic; antilipemic; cerebroprotective;  
 KW antiobese; osteoporotic; antidiabetic; cytoskeletal; antidepressant;  
 KW immunomodulator; antianemic; tranquilizer; antiparkinsonian; nootropic;  
 KW neuroprotective; antiinflammatory; antidiabetic; analgesic;  
 KW human citron rho/rac-interacting kinase; enzyme; CR1K; ameliorating;  
 KW obesity; comorbidities; cancer; anorexia; cachexia; bulimia;  
 KW central nervous system disorder; chronic obstructive pulmonary disease;  
 KW diabetes; pain; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003004523-A1.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 28-JUN-2002; 2002WO-EP071156.  
 XX  
 PR 02-JUL-2001; 2001US-301841P.  
 PR 11-DEC-2001; 2001US-338651P.  
 PR 25-APR-2002; 2002US-375014P.  
 XX  
 PA (FARB ) BAYER AG.  
 XX  
 PI Zhu Z;  
 XX  
 DR WPI; 2003-221576/21.  
 XX  
 PT New human citron rho/rac-interacting kinase (CR1K) polypeptide and  
 PT polynucleotide, useful in preventing, ameliorating or treating diseases  
 PT Alzheimer's disease -  
 XX  
 PS Disclosure; Page 217-222; 237pp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide encoding a human

CC citron rho/rac-interacting kinase polypeptide. The isolated  
 CC polynucleotide comprises a 6156 or 8603 base pair sequence, given in the  
 CC specification. The human citron rho/rac-interacting kinase (CR1K)  
 CC polypeptide and polynucleotide are useful in preventing, ameliorating, or  
 CC treating diseases associated with human CR1K dysfunction such as obesity  
 CC and obesity-associated comorbidities (e.g. hypertension, coronary artery  
 CC disease, hyperlipidemia, stroke, gout, osteoarthritis, some types of  
 CC cancer including endometrial, breast, prostate and colon cancer),  
 CC anorexia, cachexia, bulimia, central nervous system disorders (e.g. mood  
 CC disorders, anxiety disorders, Parkinson's disease or Alzheimer's  
 CC disease), chronic obstructive pulmonary disease, or diabetes. These can  
 CC also be used to treat pain associated with the disorders. The human CR1K  
 CC polypeptide is also useful in diagnostic assays or in genetic testing.  
 CC The expression vector or the reagent is useful in preparing a medicament  
 CC for modulating the activity of a human CR1K in a disease, e.g. obesity, a  
 CC central nervous system disorder, or chronic obstructive pulmonary  
 CC disease. The fusion protein is useful for generating antibodies against a  
 CC CR1K polypeptide and for use in various assay systems. The methods are  
 CC useful in producing and detecting the polynucleotide and polypeptide and  
 CC in screening for agents that modulate the activity of the human CR1K  
 CC polypeptide. This polynucleotide sequence represents a DNA sequence  
 CC relating to the human CR1K protein of the invention.  
 CC  
 SQ Sequence 6156 BP; 1732 A; 1548 C; 1679 G; 1197 T; 0 other;

Query Match 99.6%; Score 6142.2; DB 25; Length 6156;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 6147; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGTGAAGTTCAAAATATGAGAGCGGAATCTTTGATGCTGCTGCTGAACCCATT 60  
 Db 1 ATGTGAAGTTCAAAATATGAGAGCGGAATCTTTGATGCTGCTGCTGAACCCATT 60  
 Qy 61 GCCAGCGGGCTCCAGAGCTGAATCTGTTCCAGAGGGAACACCCCTTATGACTCAA 120  
 Db 61 GCCAGCGGGCTCCAGAGCTGAATCTGTTCCAGAGGGAACACCCCTTATGACTCAA 120  
 Qy 121 CAGCAGATGCTCTCTTCTCCGAGAGGATATAGATCCCTCTTGTCTTGGAA 180  
 Db 121 CAGCAGATGCTCTCTTCTCCGAGAGGATATAGATCCCTCTTGTCTTGGAA 180  
 Qy 181 GAATGAGTCAAGCTCTCTGATGAAGATTAGACAGCTGAGCAATTTGTCGGAAATAT 240  
 Db 181 GAATGAGTCAAGCTCTCTGATGAAGATTAGACAGCTGAGCAATTTGTCGGAAATAT 240  
 Qy 241 TCCGACACCATATGATGTTAAGAGAGCTCCAGGCTCCGCAAGAGATTCAGATCAGA 300  
 Db 241 TCCGACACCATATGATGTTAAGAGAGCTCCAGGCTCCGCAAGAGATTCAGATCAGA 300  
 Qy 301 AGCTTTGAGTGTGTGCTCACTTGTGTAAGTCAAGTGTAAAGAGAAAGCAACCGGG 360  
 Db 301 AGCTTTGAGTGTGTGCTCACTTGTGTAAGTCAAGTGTAAAGAGAAAGCAACCGGG 360  
 Qy 361 GACATCTATGATGAAATGATGAAGAAAGGCTTTATTTGGCCAGAGACAGGTTTCA 420  
 Db 361 GACATCTATGATGAAATGATGAAGAAAGGCTTTATTTGGCCAGAGACAGGTTTCA 420  
 Qy 421 TTTTGGAGAGAGCGGAAATATATTCGAAGACAAAGCCCGTGTATCCCAATTA 480  
 Db 421 TTTTGGAGAGAGCGGAAATATATTCGAAGACAAAGCCCGTGTATCCCAATTA 480  
 Qy 481 CAGTATGCTTTCAAGACAAAATCACTTTATCTGATGATGAAATATAGCTGAGAGG 540  
 Db 481 CAGTATGCTTTCAAGACAAAATCACTTTATCTGATGATGAAATATAGCTGAGAGG 540  
 Qy 541 GACTTGCCTGCTTTTGAATATGATGAGAGCCGTTTATGATGAAAACCTGATACAGTTT 600  
 Db 541 GACTTGCCTGCTTTTGAATATGATGAGAGCCGTTTATGATGAAAACCTGATACAGTTT 600  
 Qy 601 TACCTAGCTGAGCTATTTGCTGTTCAGACAGCTTCATCTGATGGATGATGATGATGCA 660  
 Db 601 TACCTAGCTGAGCTATTTGCTGTTCAGACAGCTTCATCTGATGGATGATGATGATGCA 660



Db 2821 GCCCTGAGAGCCAGCTTTCGCGCAGAGCAGAGAGCTGGAAGAGACACAGAGAGAGCT 2880  
Qy 2881 GAAAGAGAGATCCAGGACCTCAAGGCACTATAGAGATGAATCCAGCGGCAATTTGATGCT 2940  
Db 2881 GAAAGAGAGATCCAGGACCTCAAGGCACTATAGAGATGAATCCAGCGGCAATTTGATGCT 2940  
Qy 2941 CTTGCTTAACAGCTGTACTGTATCAAGACTGTGAGAGAGAGCTTAAACAGCTGACCGAG 3000  
Db 2941 CTTGCTTAACAGCTGTACTGTATCAAGACTGTGAGAGAGAGCTTAAACAGCTGACCGAG 3000  
Qy 3001 GAGCAAGCTGAATCTCAACCAACCAATTTCTTCTGTCACAACTCTGATGAGGCTTCT 3060  
Db 3001 GAGCAAGCTGAATCTCAACCAACCAATTTCTTCTGTCACAACTCTGATGAGGCTTCT 3060  
Qy 3061 GAGCAAGCTGAATCTCAACCAACCAATTTCTTCTGTCACAACTCTGATGAGGCTTCT 3060  
Db 3061 GAGCAAGCTGAATCTCAACCAACCAATTTCTTCTGTCACAACTCTGATGAGGCTTCT 3060  
Qy 3121 AGGAAACGAGAGATGACAGCTTACAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
Db 3121 AGGAAACGAGAGATGACAGCTTACAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
Qy 3181 TGACACATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240  
Db 3181 TGACACATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240  
Qy 3241 AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300  
Db 3241 AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300  
Qy 3301 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360  
Db 3301 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360  
Qy 3361 GATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420  
Db 3361 GATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420  
Qy 3421 GAGATTTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480  
Db 3421 GAGATTTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480  
Qy 3481 GAGATTTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
Db 3481 GAGATTTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
Qy 3541 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600  
Db 3541 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600  
Qy 3601 TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
Db 3601 TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
Qy 3661 GAAAGCTTGAATCGAGGCTGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720  
Db 3661 GAAAGCTTGAATCGAGGCTGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720  
Qy 3721 GAAAGCTTGAATCGAGGCTGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780  
Db 3721 GAAAGCTTGAATCGAGGCTGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780  
Qy 3781 CAAACCAAACTCATTTGATTTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840  
Db 3781 CAAACCAAACTCATTTGATTTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840  
Qy 3841 GTTCTCTGCAAGTCAATGAGCTGAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900  
Db 3841 GTTCTCTGCAAGTCAATGAGCTGAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900  
Qy 3901 GAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960  
Db 3901 GAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960

Qy 3961 GCCACCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020  
Db 3961 GCCACCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020  
Qy 4021 ATGCGCATGCTCCGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080  
Db 4021 ATGCGCATGCTCCGAGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080  
Qy 4081 GCGCGCGCATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140  
Db 4081 GCGCGCGCATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140  
Qy 4141 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200  
Db 4141 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200  
Qy 4201 ACAAAGTGTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260  
Db 4201 ACAAAGTGTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260  
Qy 4261 GAATGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320  
Db 4261 GAATGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320  
Qy 4321 CCTGCTGAATATGCAACACATTTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4380  
Db 4321 CCTGCTGAATATGCAACACATTTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4380  
Qy 4381 GGTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4440  
Db 4381 GGTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4440  
Qy 4441 AGGAAATTAACAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4500  
Db 4441 AGGAAATTAACAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4500  
Qy 4501 AAAAGTCTCATTTATGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560  
Db 4501 AAAAGTCTCATTTATGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560  
Qy 4561 GAGCTGTGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620  
Db 4561 GAGCTGTGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620  
Qy 4621 GCAAAATACAGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
Db 4621 GCAAAATACAGCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
Qy 4681 ACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4740  
Db 4681 ACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4740  
Qy 4741 TGGGTACAGCCCTTGAATCAGTTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800  
Db 4741 TGGGTACAGCCCTTGAATCAGTTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800  
Qy 4801 GCTGATGCTAACTGCTTGAATCAGTTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
Db 4801 GCTGATGCTAACTGCTTGAATCAGTTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
Qy 4861 ATGAACTGACAGCTGCTTGAATCAGTTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920  
Db 4861 ATGAACTGACAGCTGCTTGAATCAGTTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920  
Qy 4921 TACGCTGATGATGCTTGAATCAGTTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4980  
Db 4921 TACGCTGATGATGCTTGAATCAGTTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4980  
Qy 4981 CAAATTTATATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5040  
Db 4981 CAAATTTATATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5040

```

QY 5041 CTGTGCTTGTGGACGGTAAGTAAGTAAGTCCCTGGCCCAAGTCCCACTTGGCTGCC 5100
DB 5041 CTGTGCTTGTGGACGGTAAGTAAGTAAGTCCCTGGCCCAAGTCCCACTTGGCTGCC 5100
QY 5101 CAGCCCGACATCTCACCCCAACATTTTGAAGCTGTCAAGGGCTGCCACTTGTGTGGGGCA 5160
DB 5101 CAGCCCGACATCTCACCCCAACATTTTGAAGCTGTCAAGGGCTGCCACTTGTGTGGGGCA 5160
QY 5161 GGCATAATTGAGAACGGGCTGTGATCTGTGACGACATGCGCAGCAAGTGTCTATCTC 5220
DB 5161 GGCATAATTGAGAACGGGCTGTGATCTGTGACGACATGCGCAGCAAGTGTCTATCTC 5220
QY 5221 CGGTACAGGAAACCTCGCAATATCTGATCCGGAAGAGATAGAGACTTAGAGCC 5280
DB 5221 CGGTACAGGAAACCTCGCAATATCTGATCCGGAAGAGATAGAGACTTAGAGCC 5280
QY 5281 TCGAGCTGTATCCACTTCAACCAATTCAGTATCTCTATTGGAACCAATTAATCTACGA 5340
DB 5281 TCGAGCTGTATCCACTTCAACCAATTCAGTATCTCTATTGGAACCAATTAATCTACGA 5340
QY 5341 ATCGACATGAGACGTAACGCTCGAGGAATTCCTGATTAAGATGACCATTCCTTGCA 5400
DB 5341 ATCGACATGAGACGTAACGCTCGAGGAATTCCTGATTAAGATGACCATTCCTTGCA 5400
QY 5401 CCTGCTGTGTGGCCCTCTTCCAAAGCTTCCCTGCTCATGCTGAGAGTGAACAGC 5460
DB 5401 CCTGCTGTGTGGCCCTCTTCCAAAGCTTCCCTGCTCATGCTGAGAGTGAACAGC 5460
QY 5461 GCAGGGCAGCAGAGAGTACTTGTGCTTCCAGATTCCTGCTCATGCTGAGAGTGAACAGC 5520
DB 5461 GCAGGGCAGCAGAGAGTACTTGTGCTTCCAGATTCCTGCTCATGCTGAGAGTGAACAGC 5520
QY 5521 TACGGAAGCTGTAGCCGCAACAGCATCTCAAGTGAAGTGTGCTTACCTTTGGCCTTGGC 5580
DB 5521 TACGGAAGCTGTAGCCGCAACAGCATCTCAAGTGAAGTGTGCTTACCTTTGGCCTTGGC 5580
QY 5581 TACAGGAAACCTATCTGTGTGACCCCACTTCACTGCTGAAGTATTTGATCCAG 5640
DB 5581 TACAGGAAACCTATCTGTGTGACCCCACTTCACTGCTGAAGTATTTGATCCAG 5640
QY 5641 GCAGCCTCTCTAGCAGGAGCCCTGCGCAGGCTACTTGAAGTCCCGAACCCGCGCTAC 5700
DB 5641 GCAGCCTCTCTAGCAGGAGCCCTGCGCAGGCTACTTGAAGTCCCGAACCCGCGCTAC 5700
QY 5701 CTGGGCTCTGCAATTTCTCTCAAGAGGATTTACTTGGCTCTCTCAATCCAGGATTAATTA 5760
DB 5701 CTGGGCTCTGCAATTTCTCTCAAGAGGATTTACTTGGCTCTCTCAATCCAGGATTAATTA 5760
QY 5761 AGGTCATTTGCTGCAAGGAAACCTCGTGAAGAGTCCGCACTGAAACCAACCGGGGC 5820
DB 5761 AGGTCATTTGCTGCAAGGAAACCTCGTGAAGAGTCCGCACTGAAACCAACCGGGGC 5820
QY 5821 CCGTCCACCTCCCGCAGAGCCCAACGAGGCGCAACCACTGACACGAGCAGCATC 5880
DB 5821 CCGTCCACCTCCCGCAGAGCCCAACGAGGCGCAACCACTGACACGAGCAGCATC 5880
QY 5881 ACCAAGGGGTGGGCTTCCAGCCCAAGCGCGCCGCAAGGCGCCAGCCACCCGAGAGCCA 5940
DB 5881 ACCAAGGGGTGGGCTTCCAGCCCAAGCGCGCCGCAAGGCGCCAGCCACCCGAGAGCCA 5940
QY 5941 AGCACCACCCACCGCTACCGGAGGGGAGCCGAGCTGCGCAGGAGGACAAAGTCTCTGGC 6000
DB 5941 AGCACCACCCACCGCTACCGGAGGGGAGCCGAGCTGCGCAGGAGGACAAAGTCTCTGGC 6000
QY 6001 CGCCCTCTGAGAGCGAGAGAGTCCCGCGCGGATACTCAGACCGCGAGAGAGCGTCC 6060
DB 6001 CGCCCTCTGAGAGCGAGAGAGTCCCGCGCGGATACTCAGACCGCGAGAGAGCGTCC 6060
QY 6061 CCCGAGAGCTGTTTGAAGAGAGAGAGGCGCGCTGCTGCGGAGCCGCTGAGAGC 6120
DB 6061 CCCGAGAGCTGTTTGAAGAGAGAGAGGCGCGCTGCTGCGGAGCCGCTGAGAGC 6120
QY 6121 CCGCTGTCCAGGTGAACAAAGTCTGGAGCCAGTC 6155

```

```

DB 6121 CCGCTGTCCAGGTGAACAAAGTGAAGGAGCGCAGCATTC 6155
RESULT 6
ID AAS06701 standard; cDNA; 6159 BP.
XX
XX AAS06701;
XX
DE 12-SEP-2001 (first entry)
XX
XX Polynucleotide sequence encoding human protein kinase #1.
XX
XX Human; protein kinase; PK; STK; cancer; cardiovascular disease;
XX metabolic disorder; immune related disease; neurological disorder;
XX neurodegenerative disorder; inflammatory disorder; infectious disease;
XX reproductive disorder; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX MO200138503-A2.
XX
XX 31-MAY-2001.
XX
XX 22-NOV-2000; 2000MO-US32085.
XX
XX 24-NOV-1999; 99US-0167482.
XX
XX (SUBE-) SUGEN INC.
XX
XX Plowman GD, Whyte D, Manning G, Sudarsanam S, Martinez R;
XX Flanagan P, Clary D;
XX
XX WPI; 2001-343950/36.
XX
XX P-PSDB; AAU03501.
XX
XX Nucleic acids encoding human kinase polypeptides, useful for preventing
XX diagnosing and/or treating e.g. cancer, immune, cardiovascular and
XX neuronal-associated diseases, and microbial infections -
XX
XX Example 1; Figure 1; 433bp; English.
XX
XX AAS06701-AAS06757 encode for novel human protein kinases #1-57. The
XX novel protein kinases have been identified as members of the tyrosine
XX or serine/threonine kinase (PTK and STK) families. The polynucleotides
XX encoding protein kinases and the polypeptides may be used in the
XX prevention, diagnosis and treatment of diseases associated with
XX inappropriate kinase expression. For example, they may be used to treat
XX cancers (especially cancers of hematopoietic origin), cardiovascular
XX disease (e.g. atherosclerosis), metabolic disorders (e.g. diabetes),
XX immune related diseases (e.g. rheumatoid arthritis), neurological
XX disorders (e.g. schizophrenia), neurodegenerative disorders (e.g.
XX Parkinson's disease), inflammatory disorders (e.g. asthma), infectious
XX disease (e.g. HIV) and reproductive disorders (e.g. infertility).
XX Additionally, polynucleotides encoding protein kinases may be
XX used for gene therapy and as DNA probes in diagnostic assays.
XX The protein kinase polypeptides may be used as antigens in the production
XX of antibodies against the protein kinases and in assays to identify
XX modulators of protein kinase expression and activity.
XX
XX Sequence 6159 BP; 1732 A; 1549 C; 1680 G; 1198 T; 0 other;
SQ
Query Match 99.4%; Score 6126; DB 22; Length 6159;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 6145; Conservative 0; Mismatches 10; Indels 3; Gaps 1;
QY 1 ATGTTGAAGTTCAATATGAGACCGGATCTTGTGATGCTGTGTAACCATTC 60
DB 1 ATGTTGAAGTTCAATATGAGACCGGATCTTGTGATGCTGTGTAACCATTC 60
QY 61 GCCAGCCGGGCTCCAGGCTTAATCTGTTCTTCCAGGGGAAACCACTTTATGATCA 120

```



Dh 61 GCCAGCGGCGCTCCAGGCTGAATCTGTTCTCCAGGGGAAACACCCTTTATGACTCAA 120  
Qy 121 CAGAGATGTCCTCTCTTCCCGAGAGGATATTAGATGCCCTCTTGTCTCTTTGAA 180  
Dh 121 CAGAGATGTCCTCTCTTCCCGAGAGGATATTAGATGCCCTCTTGTCTCTTTGAA 180  
Qy 181 GAATGACATGACCTGCTGTATGATGATTAAGCAGCTGAGCACTTTGTCCGAAA--G 237  
Dh 181 GAATGACATGACCTGCTGTATGATGATTAAGCAGCTGAGCACTTTGTCCGAAAAGTG 240  
Qy 238 TATTCGACACCATAGCTGATTAACAGAGCTCCAGCTTCGGCAAAAGACCTTCGAAATC 237  
Dh 241 TATTCGACACCATAGCTGATTAACAGAGCTCCAGCTTCGGCAAAAGACCTTCGAAATC 300  
Qy 298 AGAGCTTGTAGGTGTGCTCACTTTGCTGAAGTGAGGCTGGTAAAGAGAACACC 357  
Dh 301 AGAGCTTGTAGGTGTGCTCACTTTGCTGAAGTGAGGCTGGTAAAGAGAACACC 360  
Qy 358 GGGGACATCTATGCTTATGAAAGTATGAAGAGAGGCTTTATTTGCCCAGAGCAGTT 417  
Dh 361 GGGGACATCTATGCTTATGAAAGTATGAAGAGAGGCTTTATTTGGCCCAGAGCAGTT 420  
Qy 418 TCATTTTGAAGAAAGCGGAAACATATATCTCGAAGCACAAGCCCGTGGATCCCCCAA 477  
Dh 421 TCATTTTGAAGAAAGCGGAAACATATATCTCGAAGCACAAGCCCGTGGATCCCCCAA 480  
Qy 478 TTACAGTATGCTTTCAAGACAAATAATCACTTTATCTGCTCATGGAATTCAGCCTGGA 537  
Dh 481 TTACAGTATGCTTTCAAGACAAATAATCACTTTATCTGCTCATGGAATTCAGCCTGGA 540  
Qy 538 GGGGACCTGCTGCTCACTTTGAAATGATATGAGGACCAAGTTAGTAAACCTGATACAG 597  
Dh 541 GGGGACCTGCTGCTCACTTTGAAATGATATGAGGACCAAGTTAGTAAACCTGATACAG 600  
Qy 598 TTTTACCTAGCTGAGCTGATTTTGGCTGTTCAAGCCTGATCTGATGAGTACGTGAT 657  
Dh 601 TTTTACCTAGCTGAGCTGATTTTGGCTGTTCAAGCCTGATCTGATGAGTACGTGAT 660  
Qy 658 CGAGACATCAAGCTTGAAACATTTCTGTTGACCGCACAGAGACATCAAGCTGTGAT 717  
Dh 661 CGAGACATCAAGCTTGAAACATTTCTGTTGACCGCACAGAGACATCAAGCTGTGAT 720  
Qy 718 TTTGATCTGCGCGGAAATGAAATTCAAACAGATGATGATGCAAACTCCGATTTGG 777  
Dh 721 TTTGATCTGCGCGGAAATGAAATTCAAACAGATGATGATGCAAACTCCGATTTGG 780  
Qy 778 ACCCAGATTAATCATGCTCTCTGAAAGTGTCTGATCTGTGATGAACGCGGATGAAAAAGCACCC 837  
Dh 781 ACCCAGATTAATCATGCTCTCTGAAAGTGTCTGATCTGTGATGAACGCGGATGAAAAAGCACCC 840  
Qy 838 TAGGCGCTGACCTGTGACCTGTGGTCACTGTGGGCTGATTTGCCATATGATGATTTATGGG 897  
Dh 841 TAGGCGCTGACCTGTGACCTGTGGTCACTGTGGGCTGATTTGCCATATGATGATTTATGGG 900  
Qy 898 AGATCCCTCTCGAGAGGAACTCTGCGAGAACCTTCAATAACATTAATGAATTTCCAG 957  
Dh 901 AGATCCCTCTCGAGAGGAACTCTGCGAGAACCTTCAATAACATTAATGAATTTCCAG 960  
Qy 958 CGGTTTTTGAATTTCCAGATGACCCCAAGTGAAGGATCTTTCTGATCTGATTTCAA 1017  
Dh 961 CGGTTTTTGAATTTCCAGATGACCCCAAGTGAAGGATCTTTCTGATCTGATTTCAA 1020  
Qy 1018 AGCTTGTGTGCGGCGAGAAAGAGAGCTGAAGTTGAAGTCTTTGCTGCCATCTTTTC 1077  
Dh 1021 AGCTTGTGTGCGGCGAGAAAGAGAGCTGAAGTTGAAGTCTTTGCTGCCATCTTTTC 1080  
Qy 1078 TTTCTTAAATTTGACTGGAACAAATTCGTAATCTTCTCTCCCTTCTGTTCCACCTTC 1137  
Dh 1081 TTTCTTAAATTTGACTGGAACAAATTCGTAATCTTCTCTCCCTTCTGTTCCACCTTC 1140  
Qy 1138 AAGCTGACATGACCTCCAAATTTGATGAACAGAGAAAGAAATTCGTGGGTTTATCC 1197  
Dh 1141 AAGCTGACATGACCTCCAAATTTGATGAACAGAGAAAGAAATTCGTGGGTTTATCC 1200

Qy 1198 TCTCCGTGACAGCTGAGCCCTCAGAGCTTCTCGAGGTGAAGAACTGCGGTTTGTGGGCTTT 1257  
Dh 1201 TCTCCGTGACAGCTGAGCCCTCAGAGCTTCTCGAGGTGAAGAACTGCGGTTTGTGGGCTTT 1260  
Qy 1258 TCGTACAGCAAGCACTGAGGATTTCTTGTAGATCTGAGCTGTGTGTGCGGCTTGAGC 1317  
Dh 1261 TCGTACAGCAAGCACTGAGGATTTCTTGTAGATCTGAGCTGTGTGTGCGGCTTGAGC 1320  
Qy 1318 TCCCTCGCCAAAGACTAGCTCCATGGAAGAAACCTTCTATCAAAAGCAAGAGACTTACA 1377  
Dh 1321 TCCCTCGCCAAAGACTAGCTCCATGGAAGAAACCTTCTATCAAAAGCAAGAGACTTACA 1380  
Qy 1378 GACTCTCAGAGCAAGTGTCAACAATGAGAGCAAGAAATGACCCGTTACATCGAGAGTG 1437  
Dh 1381 GACTCTCAGAGCAAGTGTCAACAATGAGAGCAAGAAATGACCCGTTACATCGAGAGTG 1440  
Qy 1438 TCAAGGTGAGAGGCTGTGCTTATGTCAGAAAGAGGTGAGGCTGAAAGCTTCAAGCTCAG 1497  
Dh 1441 TCAAGGTGAGAGGCTGTGCTTATGTCAGAAAGAGGTGAGGCTGAAAGCTTCAAGCTCAG 1500  
Qy 1498 AGATCCCTCTGAGCAGAGCCTTGTCTACATCAAGATGCAAGATGCAAGCTTAAAGCA 1557  
Dh 1501 AGATCCCTCTGAGCAGAGCCTTGTCTACATCAAGATGCAAGATGCAAGCTTAAAGCA 1560  
Qy 1558 AGTTTGAAGCAAGCAAGATGAGGTGCTCCAGAGAGATGACAAAGCACTGACGCTTCTC 1617  
Dh 1561 AGTTTGAAGCAAGCAAGATGAGGTGCTCCAGAGAGATGACAAAGCACTGACGCTTCTC 1620  
Qy 1618 CATGATTCAGAGACAGCAAGCCGAGGCTCCAGAAATCAAAAGACAGAGATGACAGGCT 1677  
Dh 1621 CATGATTCAGAGACAGCAAGCCGAGGCTCCAGAAATCAAAAGACAGAGATGACAGGCT 1680  
Qy 1678 CAAGTGAAAGAAATGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737  
Dh 1681 CAAGTGAAAGAAATGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
Qy 1738 AGAGGATGATCTTACAGAACTGAGCTGAGAGATCTCGGCTTGTGCTGAGAGATTC 1797  
Dh 1741 AGAGGATGATCTTACAGAACTGAGCTGAGAGATCTCGGCTTGTGCTGAGAGATTC 1800  
Qy 1798 AAGCGGAAAGCGAGAAATGTCAGATTAACGTTGAAGGCTTAAGGATCAAGGGAAGGCT 1857  
Dh 1801 AAGCGGAAAGCGAGAAATGTCAGATTAACGTTGAAGGCTTAAGGATCAAGGGAAGGCT 1860  
Qy 1858 GAAGTGGAGAAATATGCGAAACTGAGAGAAATCAAATGCTGAGCAGACTCAAAATTCAG 1917  
Dh 1861 GAAGTGGAGAAATATGCGAAACTGAGAGAAATCAAATGCTGAGCAGACTCAAAATTCAG 1920  
Qy 1918 GAGCTCCAAAGAGAACTGAGAGAGCTGTAAAGCCACAGAGAGCCACCGAGCTGTG 1977  
Dh 1921 GAGCTCCAAAGAGAACTGAGAGAGCTGTAAAGCCACAGAGAGCCACCGAGCTGTG 1980  
Qy 1978 CAGAAATATCCGCAAGAGAGAGCGAGCCAGAGAGAGCTGAGAAACTGCGAGAACGA 2037  
Dh 1981 CAGAAATATCCGCAAGAGAGAGCGAGCCAGAGAGAGCTGAGAAACTGCGAGAACGA 2040  
Qy 2038 GAGGATTTCTTCTGAAGGCACTGAGAAAGAACTGTGTGAAAGCTGAGAAACGCCGCAATCT 2097  
Dh 2041 GAGGATTTCTTCTGAAGGCACTGAGAAAGAACTGTGTGAAAGCTGAGAAACGCCGCAATCT 2100  
Qy 2098 CTGAGAGCAAGAGTAAAGAGACTGAGAGCACTGAGAGCTTAAGGAAACAGACTGAAGAT 2157  
Dh 2101 CTGAGAGCAAGAGTAAAGAGACTGAGAGCACTGAGAGCTTAAGGAAACAGACTGAAGAT 2160  
Qy 2158 GACATTCAGACAAATATCCCAACAGATCCAGAGATGAGTGTATTAATTTCTGAGGCTCGAA 2217  
Dh 2161 GACATTCAGACAAATATCCCAACAGATCCAGAGATGAGTGTATTAATTTCTGAGGCTCGAA 2220  
Qy 2218 GAGAAACATCGGAGAGGCCCAAGCTCAGCCAGACCTTAAGAAAGTGCACCTGAACAGAAA 2277  
Dh 2221 GAGAAACATCGGAGAGGCCCAAGCTCAGCCAGACCTTAAGAAAGTGCACCTGAACAGAAA 2280



QY 2278 GAGCAGACTATGAGGAAAAAGATTAAAGTTTGGACATCATGATTAAGAAAGCCTGGCT 2337  
DB 2281 GAGCAGACTATGAGGAAAAAGATTAAAGTTTGGACATCATGATTAAGAAAGCCTGGCT 2340  
QY 2338 GACAAAGAGACATGAGAAACATGATGACAGACACAGAGAGAGGCCCATGAGAGGGC 2397  
DB 2341 GACAAAGAGACATGAGAAACATGATGACAGACACAGAGAGAGGCCCATGAGAGGGC 2400  
QY 2398 AAAATTCTGAGGAAACAGAGGCGATGATCAATGCTATGGAATTCAGAGATCCCTG 2457  
DB 2401 AAAATTCTGAGGAAACAGAGGCGATGATCAATGCTATGGAATTCAGAGATCCCTG 2460  
QY 2458 GAACAGAGATTGTGGAATGTCGTGAAGCCAAATAACTTGCAGCAATATGCGCTTTT 2517  
DB 2461 GAACAGAGATTGTGGAATGTCGTGAAGCCAAATAACTTGCAGCAATATGCGCTTTT 2520  
QY 2518 ACCCAAGAGACATGAGAGGCCCAAGAGAGATGATTTCTGAATCTGAGGACAGAAATT 2577  
DB 2521 ACCCAAGAGACATGAGAGGCCCAAGAGAGATGATTTCTGAATCTGAGGACAGAAATT 2580  
QY 2578 TACCTGAGACACAGGCTGAGAAATTGGAGCCAGAACCCGAAACTGAGAGAGAGCTG 2637  
DB 2581 TACCTGAGACACAGGCTGAGAAATTGGAGCCAGAACCCGAAACTGAGAGAGAGCTG 2640  
QY 2638 GAGAAGATCAGCCACCAAGACACAGTGAACAAGATCCGCTGCTGGAATCTGAGAGACAA 2697  
DB 2641 GAGAAGATCAGCCACCAAGACACAGTGAACAAGATCCGCTGCTGGAATCTGAGAGACAA 2700  
QY 2698 TTGCGGAGAGTCAGTCTAGAGACAGAGACAGAACTGAGACTAAGGCCAGCTCACA 2757  
DB 2701 TTGCGGAGAGTCAGTCTAGAGACAGAGACAGAACTGAGACTAAGGCCAGCTCACA 2760  
QY 2758 GAGCTACAGCTCTCCCTGCAAGAGGCGAGTCAAGTTGACAGCCCTGAGGCTGACAG 2817  
DB 2761 GAGCTACAGCTCTCCCTGCAAGAGGCGAGTCAAGTTGACAGCCCTGAGGCTGACAG 2820  
QY 2818 GCGGCTCTGAGAGAGCAGCTTGCCAGAGGAGAAAGAGCTGGAAGAGACCAACACAGA 2877  
DB 2821 GCGGCTCTGAGAGAGCAGCTTGCCAGAGGAGAAAGAGCTGGAAGAGACCAACACAGA 2880  
QY 2878 GCTGAAGAGAGATCCAGGACATCCGACATAGAGATGAATCCAGGCAATTTTGT 2937  
DB 2881 GCTGAAGAGAGATCCAGGACATCCGACATAGAGATGAATCCAGGCAATTTTGT 2940  
QY 2938 GCTCTTCTTAACAGCTGTACTGTATACAGACTGAGAGAGAGCTTAAACAGCTGAC 2997  
DB 2941 GCTCTTCTTAACAGCTGTACTGTATACAGACTGAGAGAGAGCTTAAACAGCTGAC 3000  
QY 2998 GAGGACAAAGCTGAATCAACAAACCAAACTTCTACTTGTCCAAACCACTGATAGGCT 3057  
DB 3001 GAGGACAAAGCTGAATCAACAAACCAAACTTCTACTTGTCCAAACCACTGATAGGCT 3060  
QY 3058 TCTGGCCGCAACGACGAGATGTATCACTGCGAAGTGAAGTGAATCTCCGCGGAG 3117  
DB 3061 TCTGGCCGCAACGACGAGATGTATCACTGCGAAGTGAAGTGAATCTCCGCGGAG 3120  
QY 3118 ATCAAGGAAAGAGATGACAGCTTACAGCCAGAAAGAAACGATGAGAGGCTCTGAAGCC 3177  
DB 3121 ATCAAGGAAAGAGATGACAGCTTACAGCCAGAAAGAAACGATGAGAGGCTCTGAAGCC 3180  
QY 3178 ACGTGACCAATGCTGAGAGAAACAGTCAATGATTGAGAGCCCTTAAACGATGAGTGT 3237  
DB 3181 ACGTGACCAATGCTGAGAGAAACAGTCAATGATTGAGAGCCCTTAAACGATGAGTGT 3240  
QY 3238 GAAAAAGAGCGGCACTGAGAGGCTGAGAGAGCCTCTGGGTGATGAGAAATCCAGTTT 3297  
DB 3241 GAAAAAGAGCGGCACTGAGAGGCTGAGAGAGCCTCTGGGTGATGAGAAATCCAGTTT 3300  
QY 3298 GAGGTGTGGGTTTGAAGCTGACAGAGATGCTGGAACCCGAGAAACGAGAGAGGCGAGA 3357  
DB 3301 GAGGTGTGGGTTTGAAGCTGACAGAGATGCTGGAACCCGAGAAACGAGAGAGGCGAGA 3360  
QY 3358 GCCGATCAGCGGATCACCGAGTCTGCCAGGTGTGAGCTGCACTGAGAGAGACACAG 3417

DB 3361 GCCGATCAGCGGATCACCGAGTCTGCCAGGTGTGAGAGCTGCACTGAGAGAGACACAG 3420  
QY 3418 GCTGAGATTTCTGCTTGTGACAGAGCTCTCAAGAGACAGAGCTGAAGGCTCGAGGCTC 3477  
DB 3421 GCTGAGATTTCTGCTTGTGACAGAGCTCTCAAGAGACAGAGCTGAAGGCTCGAGGCTC 3480  
QY 3478 TCTGACAGCTCAATGACCTGAGAGAGAGAGCTATGCTTGAATGAATGCCCCAAGC 3537  
DB 3481 TCTGACAGCTCAATGACCTGAGAGAGAGAGCTATGCTTGAATGAATGCCCCAAGC 3540  
QY 3538 TTAACAGCAGAGCTGAGAGCTGAACAGAGCTCAACAGAGCTTCTGAGAGAGCAAGCC 3597  
DB 3541 TTAACAGCAGAGCTGAGAGCTGAACAGAGCTCAACAGAGCTTCTGAGAGAGCAAGCC 3600  
QY 3598 AAATTACAGCAGAGATGAGACCTGACAGAAATCAATTTTCCGTCTGACTCAGAGACTG 3657  
DB 3601 AAATTACAGCAGAGATGAGACCTGACAGAAATCAATTTTCCGTCTGACTCAGAGACTG 3660  
QY 3658 CAAGAAGCTCTAGATCGGCTGATCTACTGAAGACAGAAAGAGTGAATTTGAGATACG 3717  
DB 3661 CAAGAAGCTCTAGATCGGCTGATCTACTGAAGACAGAAAGAGTGAATTTGAGATACG 3720  
QY 3718 CTGAAAAACATTCAGGTTCTTATTTCTCATGAAAAAGTGAATGGAAGCACTATTTCT 3777  
DB 3721 CTGAAAAACATTCAGGTTCTTATTTCTCATGAAAAAGTGAATGGAAGCACTATTTCT 3780  
QY 3778 CAACAAACCAACCTATGATTTTCTGCAAGCCAAATGGAACCACTCTAATAAGAA 3837  
DB 3781 CAACAAACCAACCTATGATTTTCTGCAAGCCAAATGGAACCACTCTAATAAGAA 3840  
QY 3838 AAGTTCTCTGCACTACAAATGAGCTGAAGCTGGGCTGGAAGAGAGAACTGCTCT 3897  
DB 3841 AAGTTCTCTGCACTACAAATGAGCTGAAGCTGGGCTGGAAGAGAGAACTGCTCT 3900  
QY 3898 GCGAGCTTGAAGAAAGCCCTTCAAGAAACCCGATGAGCTCGGCTCCGCGGAGAGAA 3957  
DB 3901 GCGAGCTTGAAGAAAGCCCTTCAAGAAACCCGATGAGCTCGGCTCCGCGGAGAGAA 3960  
QY 3958 GCTGCCACCCGAAAGCAACGAGACCAACCAACCCATCCAGCCACGCCGAGGACAG 4017  
DB 3961 GCTGCCACCCGAAAGCAACGAGACCAACCAACCCATCCAGCCACGCCGAGGACAG 4020  
QY 4018 CAGATCGCAATGCTCGGATCGGATCGGACAGAGACACAGCCAGTGGCCATGAGGCTG 4077  
DB 4021 CAGATCGCAATGCTCGGATCGGATCGGACAGAGACACAGCCAGTGGCCATGAGGCTG 4080  
QY 4078 CTGGCCCGCCATCCAGCCGAGAAAGAGTCTTCAACTCCAGAGAAATTTAGTGGCGT 4137  
DB 4081 CTGGCCCGCCATCCAGCCGAGAAAGAGTCTTCAACTCCAGAGAAATTTAGTGGCGT 4140  
QY 4138 CTTAAGGAAAGCATGACACCAATATTTCTACCGATTTCAACGTGAGACTGAACATGGCA 4197  
DB 4141 CTTAAGGAAAGCATGACACCAATATTTCTACCGATTTCAACGTGAGACTGAACATGGCA 4200  
QY 4198 GCCACAAAGTGTGTGTGTCTGATACCGTGAATTTGACGCGCAGAGCACTCAAAATGT 4257  
DB 4201 GCCACAAAGTGTGTGTGTCTGATACCGTGAATTTGACGCGCAGAGCACTCAAAATGT 4260  
QY 4258 CTGCAAAATGTCAGGTGATGTGTACCCCAAGTGTCTCACGTGCTTCCAGCAGCACTCGGC 4317  
DB 4261 CTGCAAAATGTCAGGTGATGTGTACCCCAAGTGTCTCACGTGCTTCCAGCAGCACTCGGC 4320  
QY 4318 TTGCTGTGTAATTTGACACACACTTACCGAGGCTTCTGCGCTGACAAATGAATCTCC 4377  
DB 4321 TTGCTGTGTAATTTGACACACACTTACCGAGGCTTCTGCGCTGACAAATGAATCTCC 4380  
QY 4378 CCAAGTCTCCAGACCAAGAGGCCAGCAGAGACTTGAACCTGGAAGGCTGATGAAGGTG 4437  
DB 4381 CCAAGTCTCCAGACCAAGAGGCCAGCAGAGACTTGAACCTGGAAGGCTGATGAAGGTG 4440  
QY 4438 CCAAGAAATTAACAAACGAGAGACAGAGGCTGGGACAGAAAGTCAATTTGCTTGAAGGA 4497

Db 4441 CCCAGGAATTAACAAACGAGACAGCAAGCTGAGACAGAAATGATCAATGCTCTGAGAGGA 4500  
 QY 4498 TCAAAAGTCTCATTTATGACATGATGAAGCAAGAGCTGAGACAGAGGCGGTGGAGAA 4557  
 Db 4501 TCAAAAGTCTCATTTATGACATGATGAAGCAAGAGCTGAGACAGAGGCGGTGGAGAA 4560  
 QY 4558 TTTGAGCTGTGCTTCCGACGAGGATGATCTATTCATGATGCGGTGCTTCCGAA 4617  
 Db 4561 TTTGAGCTGTGCTTCCGACGAGGATGATCTATTCATGATGCGGTGCTTCCGAA 4620  
 QY 4618 CTGCGAAATACAGCCAAAGCAGATGTCCTATCATCATGAGATGGAATCTCACCCGAC 4677  
 Db 4621 CTGCGAAATACAGCCAAAGCAGATGTCCTATCATCATGAGATGGAATCTCACCCGAC 4680  
 QY 4678 ACCACCTGCTGGCCCGGAGAAACCTCTACTGTGAGTCCAGCTTCCCTGACAAACAG 4737  
 Db 4681 ACCACCTGCTGGCCCGGAGAAACCTCTACTGTGAGTCCAGCTTCCCTGACAAACAG 4740  
 QY 4738 GCGTGGGTCAACCGCTTGAATCAATGATGTCGAGGTGGAGAGTTTCTAGGAAACAA 4797  
 Db 4741 GCGTGGGTCAACCGCTTGAATCAATGATGTCGAGGTGGAGAGTTTCTAGGAAACAA 4800  
 QY 4798 GAAGCTGATGCTAACTGCTTGAAGAACTCCGTCGAGAACTGGAAGGTGATGACCTCTA 4857  
 Db 4801 GAAGCTGATGCTAACTGCTTGAAGAACTCCGTCGAGAACTGGAAGGTGATGACCTCTA 4860  
 QY 4858 GACATGAACCTGACGCTGCTTCACTGATGACAGAGTGTGTTGTGGGACCCGAGAGAGG 4917  
 Db 4861 GACATGAACCTGACGCTGCTTCACTGATGACAGAGTGTGTTGTGGGACCCGAGAGAGG 4920  
 QY 4918 CTCTAGCCCTGAAATGCTTGAAGAACTCCCTAACCCATGTCACAGAAATTGGAGCATC 4977  
 Db 4921 CTCTAGCCCTGAAATGCTTGAAGAACTCCCTAACCCATGTCACAGAAATTGGAGCATC 4980  
 QY 4978 TTTCAAAATTTATATATCAAGAGCCTGGAAGAGTACTCATGATAGCAGAGAAAGCGG 5037  
 Db 4981 TTTCAAAATTTATATATCAAGAGCCTGGAAGAGTACTCATGATAGCAGAGAAAGCGG 5040  
 QY 5038 GACAGTGTCTTGTGGAAGCTGGAAGAAAGTGAAGAGTCTCTGCGCAATCCCACTGCT 5097  
 Db 5041 GACAGTGTCTTGTGGAAGCTGGAAGAAAGTGAAGAGTCTCTGCGCAATCCCACTGCT 5100  
 QY 5098 GCCCAGCCGACATCTCAACCCAACTTTTGAAGGTGCAAGGCGTGCCTTGTGGG 5157  
 Db 5101 GCCCAGCCGACATCTCAACCCAACTTTTGAAGGTGCAAGGCGTGCCTTGTGGG 5160  
 QY 5158 GCAAGCAAGATTGAGAAAGGAGCTGTCATCTGTGACGCGCATGCCAGCAAAAGTCAT 5217  
 Db 5161 GCAAGCAAGATTGAGAAAGGAGCTGTCATCTGTGACGCGCATGCCAGCAAAAGTCAT 5220  
 QY 5218 CTCCGCTACAGGAAACCTCAGCAAAATAGTGCATCCGGAAGAGATGAGACCTCAGAG 5277  
 Db 5221 CTCCGCTACAGGAAACCTCAGCAAAATAGTGCATCCGGAAGAGATGAGACCTCAGAG 5280  
 QY 5278 CCTCAGCTGTATCCACTTCAACCAATTACAGTATCTCATATGAAACCAATTAATCTAC 5337  
 Db 5281 CCTCAGCTGTATCCACTTCAACCAATTACAGTATCTCATATGAAACCAATTAATCTAC 5340  
 QY 5338 GAAATCGACATGAAACCAATGACAGCTGAGAGAAATCTCTGATTAAGATGACATCTCTG 5397  
 Db 5341 GAAATCGACATGAAACCAATGACAGCTGAGAGAAATCTCTGATTAAGATGACATCTCTG 5400  
 QY 5398 GCACTGTGTGTGTTGCGCGCTTTCACACAGCTTCCCTGTCTCATAGTGCAGAGTAAAC 5457  
 Db 5401 GCACTGTGTGTGTTGCGCGCTTTCACACAGCTTCCCTGTCTCATAGTGCAGAGTAAAC 5460  
 QY 5458 AGCGAGGAGCAGGAGAGAGATCTTGTGTTTCCACGAATTTGAGTGTGTTGAT 5517  
 Db 5461 AGCGAGGAGCAGGAGAGAGATCTTGTGTTTCCACGAATTTGAGTGTGTTGAT 5520  
 QY 5518 TCTTACGAGAACGTAGCCGACAGACGATCTCAAGTGAATGCTTCTTGGCCCTT 5577  
 Db 5521 TCTTACGAGAACGTAGCCGACAGACGATCTCAAGTGAATGCTTCTTGGCCCTT 5580

QY 5578 GCCTACAGAGAAACCTATCTGTTTGTGACCACTTCACTCACTCGAAGTAATTGATC 5637  
 Db 5581 GCCTACAGAGAAACCTATCTGTTTGTGACCACTTCACTCACTCGAAGTAATTGATC 5640  
 QY 5638 CAGACAGCTCTCAGACAGAGAACCCCTGCCCCAGGCTGACATGACATCCCGAG 5697  
 Db 5641 CAGACAGCTCTCAGACAGAGAACCCCTGCCCCAGGCTGACATGACATCCCGAG 5700  
 QY 5698 TACTGGGCGCTGCGCAATTTCTCTGAGAGAGATTTACTTGGCGCTCTCATCCAGATTA 5757  
 Db 5701 TACTGGGCGCTGCGCAATTTCTCTGAGAGAGATTTACTTGGCGCTCTCATCCAGATTA 5760  
 QY 5758 TTAAGGATCATTTGTCTGCAAGGAGAACTCTGTGAAGAGTCCGAGCACTGAACACCGG 5817  
 Db 5761 TTAAGGATCATTTGTCTGCAAGGAGAACTCTGTGAAGAGTCCGAGCACTGAACACCGG 5820  
 QY 5818 GGCCTGTCACCTCCGACAGACCTCCCAACAGAGGCGCCCACTGACAGAGCAC 5877  
 Db 5821 GGCCTGTCACCTCCGACAGACCTCCCAACAGAGGCGCCCACTGACAGAGCAC 5880  
 QY 5878 ATCACCAAGGCGGTGCGCTCCAGCCAGGCGCGCCGAAAGGCCCAAGCCGAGAG 5937  
 Db 5881 ATCACCAAGGCGGTGCGCTCCAGCCAGGCGCGCCGAAAGGCCCAAGCCGAGAG 5940  
 QY 5938 CCAAGCAACCCCAACCGCTTACCGAGAGGCGGAGCCGAGCTGCGCAAGAGTCTCT 5997  
 Db 5941 CCAAGCAACCCCAACCGCTTACCGAGAGGCGGAGCCGAGCTGCGCAAGAGTCTCT 6000  
 QY 5998 GGCCTGCTTGAAGAGAGAGATGTCCTCCGCGCGGATCTAGACACCGGAGAGCGG 6057  
 Db 6001 GGCCTGCTTGAAGAGAGAGATGTCCTCCGCGCGGATCTAGACACCGGAGAGCGG 6060  
 QY 6058 TCCCGCGAGAGCTGTTTGAAGACAGACAGAGGCGCGCTGCTGCGAGCGGTGAGG 6117  
 Db 6061 TCCCGCGAGAGCTGTTTGAAGACAGACAGAGGCGCGCTGCTGCGAGCGGTGAGG 6120  
 QY 6118 ACCCGCTGTCCAGGTGAACCAAGGTCTGGACCACTC 6155  
 Db 6121 ACCCGCTGTCCAGGTGAACCAAGGTCTGGACCACTTC 6158

RESULT 7  
 ABS63436  
 ID ABS63436 standard; cDNA; 6189 BP.  
 XX  
 AC ABS63436;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE RHO/RAC-interacting citron kinase-like human cDNA, designated NOV3b.  
 XX  
 KW Human; gene; ss; NOV; NOVX-associated disorder; cardiomyopathy;  
 KW atherosclerosis; diabetes; cell signalling; metabolic pathway;  
 KW cellular receptor; downstream effector; cancer; gene therapy;  
 KW hypertension; congenital heart defect; aortic stenosis; obesity;  
 KW infectious disease; anorexia; Alzheimer's disease; Parkinson's disease;  
 KW neurodegenerative disorder; haemophilia; dyslipidemia; vaccine;  
 KW haematopoietic disease; scleroderma; fertility; immunogen;  
 KW idiopathic thrombocytopenic purpura; graft versus host disease;  
 KW Crohn's disease; multiple sclerosis; cirrhosis; autoimmune disease;  
 KW systemic lupus erythematosus; asthma; arthritis; psoriasis; allergy;  
 KW stroke; anxiety; Lesch-Nyhan syndrome; schizophrenia; cerebellar ataxia;  
 KW pain; alcoholism; transgenic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200226826-A2.  
 XX  
 PD 04-APR-2002.  
 XX  
 PF 27-SEP-2001; 2001MC-US42336.  
 XX



D 1018 TTGTTGCGGCGAAGAAAGAGAGACTGAAGTTTGAAAGTCTTGTGCGCATCCTTCTTC 1077  
Q 1081 TCTAAATGATGGAACCAATTCGTAATCTCTCCCTCCCTGCTGCTCCACCCCTCAAG 1140  
D 1078 TCTAAATGATGGAACCAATTCGTAATCTCTCCCTCCCTGCTGCTCCACCCCTCAAG 1137  
Q 1141 TCTGAGATGACACTTCCAAATTTGATGAACAAGAAATTCGTTGCTTCTCTCT 1200  
D 1138 TCTGAGATGACACTTCCAAATTTGATGAACAAGAAATTCGTTGCTTCTCTCT 1197  
Q 1201 CCGTGCAGCTGAGCCCTCAGAGCTTCTCGGTTGAAGAACTGCCGTTTGTGGGGTTTTCG 1260  
D 1198 CCGTGCAGCTGAGCCCTCAGAGCTTCTCGGTTGAAGAACTGCCGTTTGTGGGGTTTTCG 1257  
Q 1261 TACAGCAAGGCACTGAGGATTTCTTGTAGATCTGAGCTGTGTGTGCGGCTCGAGCTCC 1320  
D 1258 TACAGCAAGGCACTGAGGATTTCTTGTAGATCTGAGCTGTGTGTGCGGCTCGAGCTCC 1317  
Q 1321 CTTGCCAAGACTAGCTCCATGGAAGAACTTCTCATCAAAAGCAAAAGCTTACAAAGC 1380  
D 1318 CTTGCCAAGACTAGCTCCATGGAAGAACTTCTCATCAAAAGCAAAAGCTTACAAAGC 1377  
Q 1381 TCTCAGGCAAGTGTCAAGATGAGCAGAAATGACCCGTTTATCATGCGAGAGTGTCA 1440  
D 1378 TCTCAGGCAAGTGTCAAGATGAGCAGAAATGACCCGTTTATCATGCGAGAGTGTCA 1437  
Q 1441 GAGGTGAGGCTGTGCTTACTGTAAGAGAGGTGAGCTGAAAGCTTCTGAGCTCAGAGA 1500  
D 1438 GAGGTGAGGCTGTGCTTACTGTAAGAGAGGTGAGCTGAAAGCTTCTGAGCTCAGAGA 1497  
Q 1501 TCCCTCTGAGCAGAGCCTTGTCTACTCATCAAGAAATGCTGATGCTTAAAGCGAAGT 1560  
D 1498 TCCCTCTGAGCAGAGCCTTGTCTACTCATCAAGAAATGCTGATGCTTAAAGCGAAGT 1557  
Q 1561 TTGAGCAAGCAAGTGTGAGAGGTGTCCAGAGATGACAAAGCACTGCAAGCTTCTCAT 1620  
D 1558 TTGAGCAAGCAAGTGTGAGAGGTGTCCAGAGATGACAAAGCACTGCAAGCTTCTCAT 1617  
Q 1621 GATTCAGAGCAGAGCCGGAAGCTTCAAGAAATCAAAAGCAGAGTACCAAGCTCAA 1680  
D 1618 GATTCAGAGCAGAGCCGGAAGCTTCAAGAAATCAAAAGCAGAGTACCAAGCTCAA 1677  
Q 1681 GTGGAAGAAATGAGTGTGATGATGAATCAATTTGAAAGAGATCTTGTCTCAGCAAGAG 1740  
D 1678 GTGGAAGAAATGAGTGTGATGATGAATCAATTTGAAAGAGATCTTGTCTCAGCAAGAG 1737  
Q 1741 CGAGTGTCTTACGAATGTGAGCTGAGAGAGTCTCGAGCTTGTCTGTAAGAAATTCAG 1800  
D 1738 CGAGTGTCTTACGAATGTGAGCTGAGAGAGTCTCGAGCTTGTCTGTAAGAAATTCAG 1797  
Q 1801 CGGAAAGCAGAGATGTCAAGCTTAACTGTGAAGCTTAAAGATCAAGGAAAGCTTGA 1860  
D 1798 CGGAAAGCAGAGATGTCAAGCTTAACTGTGAAGCTTAAAGATCAAGGAAAGCTTGA 1857  
Q 1861 GTGGAAGAAATGAGAACTGAGAAATCAATGTCTGAGCAGAGCTCAAAATTCAGAG 1920  
D 1858 GTGGAAGAAATGAGAACTGAGAAATCAATGTCTGAGCAGAGCTCAAAATTCAGAG 1917  
Q 1921 CTTCAAGAGAACTGAGAGAGGCTTAAAGCCAGAGCCAGAGCTGAGCTGAG 1980  
D 1918 CTTCAAGAGAACTGAGAGAGGCTTAAAGCCAGAGCCAGAGCTGAGCTGAG 1977  
Q 1981 AATATCCGCAAGCAAGAGCAGAGCCAGAGAGGAGTGTGAAGCTGCAAGACCGAGAG 2040  
D 1978 AATATCCGCAAGCAAGAGCAGAGCCAGAGAGGAGTGTGAAGCTGCAAGACCGAGAG 2037  
Q 2041 GATTTCTTGAAGGATCAGAAAGAGCTGTGAAAGCTGAGAGCGCCGCTTCTCTG 2100  
D 2038 GATTTCTTGAAGGATCAGAAAGAGCTGTGAAAGCTGAGAGCGCCGCTTCTCTG 2097  
Q 2101 GAGAAACAAGTAAAGAGACTAGAGCAGTGAAGCTTGAAGAAACAAGTGAAGATGAG 2160  
D 2098 GAGAAACAAGTAAAGAGACTAGAGCAGTGAAGCTTGAAGAAACAAGTGAAGATGAG 2157

Q 2161 ATCCAGCAAAATCCCAACAGATCCAGAGAGTGTATAAATTTCTGAGCTGAAAG 2220  
D 2158 ATCCAGCAAAATCCCAACAGATCCAGAGAGTGTATAAATTTCTGAGCTGAAAG 2217  
Q 2221 AAACATCGGAGGCCCAAGCTCTCAGCCAGAGACCTTGAAGTGCACCTGAAACAGAAAG 2280  
D 2218 AAACATCGGAGGCCCAAGCTCTCAGCCAGAGACCTTGAAGTGCACCTGAAACAGAAAG 2277  
Q 2281 CAGCACTGTGAGGAAATTAAGTGTGGAACAATCAATGAATTAAGAAAGACTGCTGAC 2340  
D 2278 CAGCACTGTGAGGAAATTAAGTGTGGAACAATCAATGAATTAAGAAAGACTGCTGAC 2337  
Q 2341 AAGAGACACTGAGAAACATGATGAGAGACAGAGAGAGGCCCATGAAAGGCAAA 2400  
D 2338 AAGAGACACTGAGAAACATGATGAGAGACAGAGAGAGGCCCATGAAAGGCAAA 2397  
Q 2401 ATTCTCAGGAAACAGAGGCGATGATCAATGTATGATTTCCAGATCAGATCCTGAA 2460  
D 2398 ATTCTCAGGAAACAGAGGCGATGATCAATGTATGATTTCCAGATCAGATCCTGAA 2457  
Q 2461 CAGAGATTGTGGAATCTGTGAAGCAATTAACCTTGACGAAATAGCAGCTTTTTC 2520  
D 2458 CAGAGATTGTGGAATCTGTGAAGCAATTAACCTTGACGAAATAGCAGCTTTTTC 2517  
Q 2521 CAAAGGAAATGAAAGGCCCAAGAGATGATTTCTGAATCTCAGGCAACAGAAATTTTAC 2580  
D 2518 CAAAGGAAATGAAAGGCCCAAGAGATGATTTCTGAATCTCAGGCAACAGAAATTTTAC 2577  
Q 2581 CTGAGAACACAGGCTGTGGAAGTTGGAAGGCCCAAGAACTGGAAGACACAGTGGAG 2640  
D 2578 CTGAGAACACAGGCTGTGGAAGTTGGAAGGCCCAAGAACTGGAAGACACAGTGGAG 2637  
Q 2641 AAGATCAAGCCCAAGAACCAAGTGAACAAGATCGGCTGTGGAATCTGGAACAAGATTG 2700  
D 2638 AAGATCAAGCCCAAGAACCAAGTGAACAAGATCGGCTGTGGAATCTGGAACAAGATTG 2697  
Q 2701 CGGAGGTCTAGTCTTGAAGCAGAGAGCAAGAACTGAGCTCAAGCCGACTCAGAG 2760  
D 2698 CGGAGGTCTAGTCTTGAAGCAGAGAGCAAGAACTGAGCTCAAGCCGACTCAGAG 2757  
Q 2761 CTACAGCTCTCCCTCAGAGAGGCGAGTCAAGATTGACAGCCCTGCAAGCTGCAAGGCG 2820  
D 2758 CTACAGCTCTCCCTCAGAGAGGCGAGTCAAGATTGACAGCCCTGCAAGCTGCAAGGCG 2817  
Q 2821 GCCCTGAGAGGCCAGCTTCCGCAAGCGAAGCAAGACTGGAAGAGCCACAGCAGAGCT 2880  
D 2818 GCCCTGAGAGGCCAGCTTCCGCAAGCGAAGCAAGACTGGAAGAGCCACAGCAGAGCT 2877  
Q 2881 GAAAGAGGATCTCAGAGCACTCAGGCACTAGAGATGAATCCAGCGCAATTTGATGCT 2940  
D 2878 GAAAGAGGATCTCAGAGCACTCAGGCACTAGAGATGAATCCAGCGCAATTTGATGCT 2937  
Q 2941 CTTCTTAACAGCTGACTGTATATCAAGACTGAGAGAGCACTTAAACAGACTGACCAG 3000  
D 2938 CTTCTTAACAGCTGACTGTATATCAAGACTGAGAGAGCACTTAAACAGACTGACCAG 2997  
Q 3001 GACAAAGCTGAACTCAACAAACAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3060  
D 2998 GACAAAGCTGAACTCAACAAACAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3057  
Q 3061 GCGCGCAAGCAGAGATGTCAACTGAGAGTGAAGTGAACATCTCCGCGGAGAGATC 3120  
D 3058 GCGCGCAAGCAGAGATGTCAACTGAGAGTGAAGTGAACATCTCCGCGGAGAGATC 3117  
Q 3121 ACGGAACAGAGATGACCTTACCAAGCAAGCAAGATGAGAGCTTGAAGCAAG 3180  
D 3118 ACGGAACAGAGATGACCTTACCAAGCAAGCAAGATGAGAGCTTGAAGCAAG 3177  
Q 3181 TGCAACATGCTGAGAGAAACAGTATGATTTGAGAGGCCCTTAAACAGATGAGCTCTAGAA 3240  
D 3178 TGCAACATGCTGAGAGAAACAGTATGATTTGAGAGGCCCTTAAACAGATGAGCTCTAGAA 3237

[illegible]

Db	4316	CTGCGTAATTATGCCACACACTTCAACCGAGGCTTCTCGCGTGAACAAATGAACCTCCCA	4377	
QY	4381	GGTCTCCAGACCAAGAGCCGACGACAGCTTGCACCTTGAAGGTGATGAAGTGTCC	4440	
Db	4378	GGTCTCCAGACCAAGAGCCGACGACAGCTTGCACCTTGAAGGTGATGAAGTGTCC	4437	
QY	4441	AGGAATAAATAACGAGACACGTAAGCTGGGACAGAAAGTACATTGTCTTGAGAGATCA	4500	
Db	4438	AGGAATAAATAACGAGACACGTAAGCTGGGACAGAAAGTACATTGTCTTGAGAGATCA	4497	
QY	4501	AAAGTCTCATTTATGACATGAACCCAGAGAAGCTGGAACAGAGCCGGTGGAGAATTT	4566	
Db	4498	AAAGTCTCATTTATGACATGAACCCAGAGAAGCTGGAACAGAGCCGGTGGAGAATTT	4557	
QY	4561	GAGCTGTGCTTCCCGACGGGATGTATCTATTCATGCTGCGTGTGCTTCCGAATC	4620	
Db	4558	GAGCTGTGCTTCCCGACGGGATGTATCTATTCATGCTGCGTGTGCTTCCGAATC	4617	
QY	4621	GGAATATACAGCCAAAGACAGATGTCCATACATCTGAGAGATGAAATCTCACCCGACAC	4680	
Db	4618	GGAATATACAGCCAAAGACAGATGTCCATACATCTGAGAGATGAAATCTCACCCGACAC	4677	
QY	4681	ACCTGCTGAGCCGGAGAACCTCTACTGTGTAGCTCCAGCTTCCCTGACAAACAGGCG	4740	
Db	4678	ACCTGCTGAGCCGGAGAACCTCTACTGTGTAGCTCCAGCTTCCCTGACAAACAGGCG	4737	
QY	4741	TGGGTACCGGCTTGAATTCAGTGTGTGCAAGTGGGAGAGTTTCTAGGAAAAAGCGAA	4800	
Db	4738	TGGGTACCGGCTTGAATTCAGTGTGTGCAAGTGGGAGAGTTTCTAGGAAAAAGCGAA	4797	
QY	4801	GCTGATGCTTAACTGCTTGAAATCTCCCTGTGTGAATCTGAAAGGTGATGACCGCTAGAC	4860	
Db	4798	GCTGATGCTTAACTGCTTGAAATCTCCCTGTGTGAATCTGAAAGGTGATGACCGCTAGAC	4857	
QY	4861	ATGAATCTGACGCTGCTCTTCACTGATCAGATGATGTGTGTGGCACCGAGAGAGGCTC	4920	
Db	4858	ATGAATCTGACGCTGCTCTTCACTGATCAGATGATGTGTGTGGCACCGAGAGAGGCTC	4917	
QY	4921	TACGCCCTGAAATGTCTTGAAAAATCCCTTAACCAATGCCACAGAAATTGAGACAGTTTC	4980	
Db	4918	TACGCCCTGAAATGTCTTGAAAAATCCCTTAACCAATGCCACAGAAATTGAGACAGTTTC	4977	
QY	4981	CAAAATTATATTAATCAAGACCTGAGAGAGTACTCATGATAGAGGTGAAGAGCGGCA	5040	
Db	4978	CAAAATTATATTAATCAAGACCTGAGAGAGTACTCATGATAGAGGTGAAGAGCGGCA	5037	
QY	5041	CTGATGCTTGAGACGTGAAGAAAGTGAACAGTCTCTGGCCAGATGCCACTGCTGCC	5100	
Db	5038	CTGATGCTTGAGACGTGAAGAAAGTGAACAGTCTCTGGCCAGATGCCACTGCTGCC	5097	
QY	5101	CAGCCCGACATCTACCCCAACATTTTGAAGCTGTCAAGGGCTGCCACTTGTGGGGCA	5160	
Db	5098	CAGCCCGACATCTACCCCAACATTTTGAAGCTGTCAAGGGCTGCCACTTGTGGGGCA	5157	
QY	5161	GGCAAGATTTGAAACGGGCTCTGTGATCTGTGACAGCCATGCCCAGCAAGTGGTATTCTC	5220	
Db	5158	GGCAAGATTTGAAACGGGCTCTGTGATCTGTGACAGCCATGCCCAGCAAGTGGTATTCTC	5217	
QY	5221	CGCTACAAAGAAAACTCAGCAAAATCTGATATCCGAAAAAGATTAAGACCTCAGAGCC	5280	
Db	5218	CGCTACAAAGAAAACTCAGCAAAATCTGATATCCGAAAAAGATTAAGACCTCAGAGCC	5277	
QY	5281	TGCAAGCTGTATTCACATTCACCAATTACAGTATCTCATTTGAAACCAATAATTCTACGAA	5340	
Db	5278	TGCAAGCTGTATTCACATTCACCAATTACAGTATCTCATTTGAAACCAATAATTCTACGAA	5337	
QY	5341	ATGCAACATGAAGAGATACAGCTCGAGGAATTTCTGGATTAAGATGACCAATTCCTTGGCA	5400	
Db	5338	ATGCAACATGAAGAGATACAGCTCGAGGAATTTCTGGATTAAGATGACCAATTCCTTGGCA	5397	
QY	5401	CTGCTGTGTGTTGCGCGCTCTTCCAAACAGCTTCCCTGTCTCAATCGTGAAGGTGAACAGC	5460	



Db 5398 COTGCTGTTGGCGGCTCTTCAGACAGCTCCCTGCTCAATCGTCAGAGTGAACAGC 5457  
 Oy 5461 GAGGAGGACGACGAGAGTACTTCTGTTGTTCCAGAAATTGGAGTGTCTGAGATCTT 5520  
 Db 5458 GAGGAGGACGACGAGAGTACTTCTGTTGTTCCAGAAATTGGAGTGTCTGAGATCTT 5517  
 Oy 5521 TACGAGAGAGTACGACGACGAGATCTCAAGAGAGTGTCTGAGTGTCTGAGTGTCTG 5580  
 Db 5518 TACGAGAGAGTACGACGACGAGATCTCAAGAGAGTGTCTGAGTGTCTGAGTGTCTG 5577  
 Oy 5581 TACGAGAGAGTACGACGACGAGATCTCAAGAGAGTGTCTGAGTGTCTGAGTGTCTG 5640  
 Db 5578 TACGAGAGAGTACGACGACGAGATCTCAAGAGAGTGTCTGAGTGTCTGAGTGTCTG 5637  
 Oy 5641 GAGGAGGACGACGAGAGTACTTCTGTTGTTCCAGAAATTGGAGTGTCTGAGATCTT 5700  
 Db 5638 GAGGAGGACGACGAGAGTACTTCTGTTGTTCCAGAAATTGGAGTGTCTGAGATCTT 5697  
 Oy 5701 CTGGGAGGACGACGAGAGTACTTCTGTTGTTCCAGAAATTGGAGTGTCTGAGATCTT 5760  
 Db 5698 CTGGGAGGACGACGAGAGTACTTCTGTTGTTCCAGAAATTGGAGTGTCTGAGATCTT 5757  
 Oy 5761 AGGGTCAATTTCTGACGAGAGAACTCTGTGAAGAGTGTCTGAGTGTCTGAGTGTCTG 5820  
 Db 5758 AGGGTCAATTTCTGACGAGAGAACTCTGTGAAGAGTGTCTGAGTGTCTGAGTGTCTG 5817  
 Oy 5821 CCGTCAACCTCCGACGAGAGAGCCCAAGAGAGAGCCCAAGTCAACGAGAGATC 5880  
 Db 5818 CCGTCAACCTCCGACGAGAGAGCCCAAGAGAGAGCCCAAGTCAACGAGAGATC 5877  
 Oy 5881 ACCAAGGAGGCTGAGGCTCCAGAGAGAGCCCAAGAGAGAGCCCAAGTCAACGAGAGATC 5940  
 Db 5878 ACCAAGGAGGCTGAGGCTCCAGAGAGAGCCCAAGAGAGAGCCCAAGTCAACGAGAGATC 5937  
 Oy 5941 AGCAACACCCGACGAGAGAGAGCCCAAGAGAGAGCCCAAGTCAACGAGAGATC 6000  
 Db 5938 AGCAACACCCGACGAGAGAGAGCCCAAGAGAGAGCCCAAGTCAACGAGAGATC 5997  
 Oy 6001 CGCCCCCTGAGAGAGAGAGTCCCGGCGGAGTACTCAGACAGCGAGAGAGAGCGTCC 6060  
 Db 5998 CGCCCCCTGAGAGAGAGAGTCCCGGCGGAGTACTCAGACAGCGAGAGAGAGCGTCC 6057  
 Oy 6061 CCCCCGAGGCTGTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6120  
 Db 6058 CCCCCGAGGCTGTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6117  
 Oy 6121 CCCCCGAGGCTGTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6165  
 Db 6118 CCCCCGAGGCTGTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6162  
 RESULT 8  
 ABS63435  
 ID ABS63435 standard; cDNA; 6201 BP.  
 AC ABS63435;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Human cDNA, homologous to kinases, designated NOV3a.  
 XX  
 KW Human; gene; ss; NOV; NOVX; NOVX-associated disorder; cardiomyopathy;  
 KW atherosclerosis; diabetes; cell signalling; metabolic pathway;  
 KW cellular receptor; downstream effector; cancer; gene therapy;  
 KW hypertension; congenital heart defect; aortic stenosis; obesity;  
 KW infectious disease; anorexia; Alzheimer's disease; Parkinson's disease;  
 KW neurodegenerative disorder; haemophilia; dyslipidemia; vaccine;  
 KW haematopoietic disease; scleroderma; fertility; immunogen;  
 KW idiopathic thrombocytopenic purpura; graft versus host disease;  
 KW Crohn's disease; multiple sclerosis; cirrhosis; autoimmune disease;  
 KW systemic lupus erythematosus; asthma; arthritis; psoriasis; allergy;  
 KW stroke; anxiety; Lesch-Nyhan syndrome; schizophrenia; cerebellar ataxia;  
 KW pain; alcoholism; transgenic.

XX Homo sapiens.  
 OS  
 XX WO200226826-A2.  
 PN  
 XX  
 PD 04-APR-2002.  
 XX  
 PF 27-SEP-2001; 2001WO-US42336.  
 XX  
 PR 27-SEP-2000; 2000US-235631P.  
 XX  
 PR 27-SEP-2000; 2000US-235633P.  
 PR 27-SEP-2000; 2000US-235808P.  
 PR 27-SEP-2000; 2000US-236064P.  
 PR 27-SEP-2000; 2000US-236065P.  
 PR 27-SEP-2000; 2000US-236066P.  
 PR 28-SEP-2000; 2000US-236135P.  
 PR 03-OCT-2000; 2000US-237434P.  
 PR 05-OCT-2000; 2000US-238321P.  
 PR 06-OCT-2000; 2000US-238396P.  
 PR 16-MAR-2001; 2001US-276667P.  
 PR 31-MAY-2001; 2001US-294823P.  
 PR 12-JUL-2001; 2001US-304868P.  
 PR 26-SEP-2001; 2001US-0235631.  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Gerlach VL, Macdougall JR, Smithson G, Millet I, Stone D;  
 PI Gunther E, Ellerman K, Grosse KM, Alsbrook JP, Lepley DM;  
 PI Burgess CE, Padigaru M, Kexuda R, Spyrek KA, Leach WD;  
 PI Shinkens RA;  
 XX  
 DR WPI; 2002-499860/53.  
 XX  
 PT P-PDB; ABG78362.  
 DR  
 XX  
 PT Novel isolated NOVX polypeptides and polynucleotides homologous to  
 PT attractin, plexin, papin-like family of proteins, useful for treating  
 PT atherosclerosis, diabetes, cancer, Alzheimer's disease, hemophilia and  
 PT stroke  
 XX  
 PS Claim 8; Page 40-42; 308pp; English.  
 XX  
 CC The invention discloses the isolated human polypeptides, and  
 CC polynucleotides encoding them, that have been designated NOVX. The  
 CC polypeptides, polynucleotides and antibodies are useful in treating or  
 CC preventing a NOVX-associated disorder which is cardiomyopathy,  
 CC atherosclerosis and diabetes in a human, where the disorder is related to  
 CC cell signal processing and metabolic pathway modulation. They can also be  
 CC used in determining the presence of, or predisposition to, a disease  
 CC associated with altered levels of the polypeptides and polynucleotides of  
 CC any one of the 13 sequences (NOV1-NOV8), for raising antibodies, for  
 CC identifying an agent that binds to, or that modulates the expression or  
 CC activity of the polypeptide, for identifying an agent which is cellular  
 CC receptor or downstream effector, for treating or preventing a  
 CC NOVX-associated disorder and as a pharmaceutical composition comprising  
 CC the polypeptide, polynucleotide or the antibody. The polypeptides and  
 CC polynucleotides are useful in diagnostic applications (e.g. as a marker  
 CC for cancerous cells or tissue types) where their amounts are assessed, or  
 CC for the manufacture of a medicament (e.g. gene therapy) for treating or  
 CC preventing disorders or syndromes such as hypertension, congenital heart  
 CC defects, aortic stenosis, obesity, infectious disease, anorexia, cancer,  
 CC Alzheimer's disease, Parkinson's disorders, neurodegenerative disorders,  
 CC haemophilia, dyslipidemia, haematopoietic diseases, scleroderma,  
 CC fertility, idiopathic thrombocytopenic purpura, graft versus host  
 CC disease, Crohn's disease, multiple sclerosis, cirrhosis, autoimmune  
 CC disease, systemic lupus erythematosus, asthma, arthritis, psoriasis,  
 CC allergy, stroke, anxiety, Lesch-Nyhan syndrome, schizophrenia,  
 CC cerebellar ataxia, pain and alcoholism. They may also be used as  
 CC immunogens to produce antibodies specific for the invention, and as  
 CC vaccines. Transgenic cells containing a NOVX expressing construct are  
 CC useful to produce non-human transgenic animals for studying the function  
 CC and/or activity of the NOVX proteins and for identifying and/or  
 CC evaluating modulators of NOVX protein activity. Transgenic cells



CC containing a NOVX expressing construct are useful to produce non-human  
CC transgenic animals for studying the function and/or activity of the NOVX  
CC proteins and for identifying and/or evaluating modulators of NOVX protein  
CC activity. The sequences presented in AB563431-AB563444 are the human  
CC NOV1-NOV8 cDNAs.

XX  
SQ Sequence 6201 BP; 1736 A; 1552 C; 1707 G; 1206 T; 0 other;

Query Match 99.0%; Score 6105.4; DB 24; Length 6201;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 6131; Conservative 0; Mismatches 21; Indels 3; Gaps 1;

```
QY 1 ATGTGAAGTTCAAAATATGAGCGCGGAATCCTTTGAGTCGTGTCTGTAACCCATT 60
DB 1 ATGTGAAGTTCAAAATATGAGCGCGGAATCCTTTGAGTCGTGTCTGTAACCCATT 60
QY 61 GCCAGCGGGCCCTCCAGGCGTGATCTGTCTCCAGGGGAACACCCCTTTAGACTCAA 120
DB 61 GCCAGCGGGCCCTCCAGGCGTGATCTGTCTCCAGGGGAACACCCCTTTAGACTCAA 120
QY 121 CAGCAGATGTCCTCTCTTCCGAGAAAGGATATAGATGCCCTCTTGTCTCTTGAA 180
DB 121 CAGCAGATGTCCTCTCTTCCGAGAAAGGATATAGATGCCCTCTTGTCTCTTGAA 180
QY 181 GAAATGAGTCAGCTGCTCTGATGAGATTAAGCAGTGAACAATTGTCCGGAATAT 240
DB 181 GAAATGAGTCAGCTGCTCTGATGAGATTAAGCAGTGAACAATTGTCCGGAATAT 240
QY 241 TCCGACACCATAGCTAGTTACAGAGAGCTCCAGCTTCCGCAAAAGACTTCGAAATCAG 300
DB 241 TCCGACACCATAGCTAGTTACAGAGAGCTCCAGCTTCCGCAAAAGACTTCGAAATCAG 300
QY 301 AGTCTTGATAGTTGTGTGTCATCTTGTCTGAAGTCAGAGTGTGAAGAAAGCAACGG 360
DB 301 AGTCTTGATAGTTGTGTGTCATCTTGTCTGAAGTCAGAGTGTGAAGAAAGCAACGG 360
QY 361 GACATCTATGCTATGAAAGTATGAAGAAAGAGCTTTATTTGCCCCAGAGAGTTTCA 420
DB 361 GACATCTATGCTATGAAAGTATGAAGAAAGAGCTTTATTTGCCCCAGAGAGTTTCA 420
QY 421 TTTTGGAGAGAGCGGAACATATTATCTCGAAGCAACAGCCCGTGGATCCCCCAATTA 480
DB 421 TTTTGGAGAGAGCGGAACATATTATCTCGAAGCAACAGCCCGTGGATCCCCCAATTA 480
QY 481 CAGTATGCTTTTCAGAGCAAAAATCACTTTATCTGTGATGGAATATCAAGCTGAGGG 540
DB 481 CAGTATGCTTTTCAGAGCAAAAATCACTTTATCTGTGATGGAATATCAAGCTGAGGG 540
QY 541 GACTTGTCTGCTTTTGAATAGATATGAGACCACTTATGATGAAATATCAAGCTGAGTTT 600
DB 541 GACTTGTCTGCTTTTGAATAGATATGAGACCACTTATGATGAAATATCAAGCTGAGTTT 600
QY 601 TACCTGAGTGAATTTTGGCTGTTCACAGGCTTCATCTGATGGGATACGTGATCGA 660
DB 601 TACCTGAGTGAATTTTGGCTGTTCACAGGCTTCATCTGATGGGATACGTGATCGA 660
QY 661 GACATCAAGCCTGAGAAATTCCTGTTGACCGGACAGACACATCAAGCTGTGATTTT 720
DB 661 GACATCAAGCCTGAGAAATTCCTGTTGACCGGACAGACACATCAAGCTGTGATTTT 720
QY 721 GGAATCGCGCGGAAATGAATTCAAAACAAGTGTGAATGCCAAATCCCGATTTGGGACC 780
DB 721 GGAATCGCGCGGAAATGAATTCAAAACAAGTGTGAATGCCAAATCCCGATTTGGGACC 780
QY 781 CCAGATTACATGCTCTGTAAGTCTGACTGTGATGAACGGGATGGAAGGACCTTAC 840
DB 781 CCAGATTACATGCTCTGTAAGTCTGACTGTGATGAACGGGATGGAAGGACCTTAC 840
QY 841 GGCCTGGAATCTGACTGTGCTGATGAGGCGGTGATTTGCTTATGAGATGATTTATGGGAGA 900
DB 841 GGCCTGGAATCTGACTGTGCTGATGAGGCGGTGATTTGCTTATGAGATGATTTATGGGAGA 900
QY 901 TCCCCCTTGCAGAGGGAACCTCTGCCAGAACCTTCAATTAACATTATGATTTCCAGCGG 960
DB 901 TCCCCCTTGCAGAGGGAACCTCTGCCAGAACCTTCAATTAACATTATGATTTCCAGCGG 960
```

```
DB 898 TCCCCCTTGCAGAGGGAACCTCTGCCAGAACCTTCAATTAACATTATGATTTCCAGCGG 957
QY 961 TTTTGAATTTTCCAAATATACCCCAAGAGAGAGACTTTTATATCTGATTCAAAGC 1020
DB 958 TTTTGAATTTTCCAAATATACCCCAAGAGAGAGACTTTTATATCTGATTCAAAGC 1017
QY 1021 TTGTGTGCGGCGCAGAAAGAGAGACTGAAGTTGAAGTCTTGTCTGCCATCTTCTTC 1080
DB 1018 TTGTGTGCGGCGCAGAAAGAGAGAGACTGAAGTTGAAGTCTTGTCTGCCATCTTCTTC 1077
QY 1081 TCTAAATTTGACTGGAACAACATTCCTCTCTCCCTCTTCTGTTCCACCTCAAG 1140
DB 1078 TCTAAATTTGACTGGAACAACATTCCTCTCTCTCCCTCTTCTGTTCCACCTCAAG 1137
QY 1141 TCTGAAGATGACACCTCCCAATTTTGAATGAACCGAAGAAATTCGTGGGTTTATCTCT 1200
DB 1138 TCTGAAGATGACACCTCCCAATTTTGAATGAACCGAAGAAATTCGTGGGTTTATCTCT 1197
QY 1201 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1260
DB 1198 CCGTGCACGCTGAGCCCTCAAGCTTCTCGAGTGAAGAACTGCGGTTGTGGGGTTTTCG 1257
QY 1261 TACAGCAAGGCACTGGGGAATCTTGTGATATCTGATCTGTGTGTGTGTGTGTGACTCC 1320
DB 1258 TACAGCAAGGCACTGGGGAATCTTGTGATATCTGATCTGTGTGTGTGTGTGTGACTCC 1317
QY 1321 CCGGCCAAGACTAGCTCCATGGAAGAAATCTTCATCAAAAGCAAAAGGCTACAGAG 1380
DB 1318 CCGGCCAAGACTAGCTCCATGGAAGAAATCTTCATCAAAAGCAAAAGGCTACAGAG 1377
QY 1381 TCTCAGAGCAAGTGTCAACAAGATGAGCAGAAATACCCGTTTATCTCGAGAGTGTCA 1440
DB 1378 TCTCAGAGCAAGTGTCAACAAGATGAGCAGAAATACCCGTTTATCTCGAGAGTGTCA 1437
QY 1441 GAGGTGAGGCTGTGCTTATGTCAGAAAGAGTGAAGTGAAGGCTTGTGACTCAGAGA 1500
DB 1438 GAGGTGAGGCTGTGCTTATGTCAGAAAGAGTGAAGTGAAGGCTTGTGACTCAGAGA 1497
QY 1501 TCCCTCTGAGAGAGAGACCTTGCTACTTCATCTCAGAAATGCAATGAGCTTAAAGCCGAA 1560
DB 1498 TCCCTCTGAGAGAGAGACCTTGCTACTTCATCTCAGAAATGCAATGAGCTTAAAGCCGAA 1557
QY 1561 TTGGAACAAGCAACGATGAGGCTGTCCAGAGAGATGACAAAGCACTGACGTTCTCCAT 1620
DB 1558 TTGGAACAAGCAACGATGAGGCTGTCCAGAGAGATGACAAAGCACTGACGTTCTCCAT 1617
QY 1621 GATATCAGAGAGAGAGCCGGAAGCTCCAGAAATCAAAAGCAGAGATCAAGGCTCAA 1680
DB 1618 GATATCAGAGAGAGAGCCGGAAGCTCCAGAAATCAAAAGCAGAGATCAAGGCTCAA 1677
QY 1681 GTGGAAGAAATGAGTGTGATGAATCAAGTTGGAAGAGGATCTTGTCTCAGCAAGAG 1740
DB 1678 GTGGAAGAAATGAGTGTGATGAATCAAGTTGGAAGAGGATCTTGTCTCAGCAAGAG 1737
QY 1741 CGGAGTATCTCTACGAATCTGAGAGAGTCTCGGCTTGTCTGTGAAGATTTCAAG 1800
DB 1738 CGGAGTATCTCTACGAATCTGAGAGAGTCTCGGCTTGTCTGTGAAGATTTCAAG 1797
QY 1801 CGGAAAGCGACAGATGTGAGCATTAACCTGTTGAAGGCTTAAGATCAAGGAAAGCCTGAA 1860
DB 1798 CGGAAAGCGACAGAAATGTGAGCATTAACCTGTTGAAGGCTTAAGATCAAGGAAAGCCTGAA 1857
QY 1861 GTGGAGAAATATGCGGAACTGGAGAAATCAATGTGAGCAGAGAGCTCAAAATTCAGAG 1920
DB 1858 GTGGAGAAATATGCGGAACTGGAGAAATCAATGTGAGCAGAGAGCTCAAAATTCAGAG 1917
QY 1921 CTCGAAGAGAAATCTGGAGAAAGCTGTAAAGCCAGGACCGAGGCTGTGAG 1980
DB 1918 CTCGAAGAGAAATCTGGAGAAAGCTGTAAAGCCAGGACCGAGGCTGTGAG 1977
QY 1981 AATATCCGCGCAGGCAAGAGAGGAGCCGAGAGAGAGCTGGAAGAGCTGCAAGAACGAGAG 2040
```

Db 1978 AATATCCGCCAGGCAAAAGAGGAGCCGAGAGGAGCTGGAGAACTGCAAGAACCGAGAG 2037  
Qy 2041 GATTCTTCTGAAGGATGAGAAAGAAAGCTGGTGAAGGCTGAGAAACGCGCCGCTTCTGTG 2100  
Db 2038 GATTCTTCTGAAGGATGAGAAAGAAAGCTGGTGAAGGCTGAGAAACGCGCCGCTTCTGTG 2097  
Qy 2101 GAGAAACAAGGTAAAGAGCTAGAGACATAGAGCTAGAGAAAACAGACTGAGAGTAGAC 2160  
Db 2098 GAGAAACAAGGTAAAGAGCTAGAGACATAGAGCTAGAGAAAACAGACTGAGAGTAGAC 2157  
Qy 2161 ATCCAGACAAAATCCCAACAGATCCAGAGATGGCTGATTAATAATTCTGAGCTCGAAGAG 2220  
Db 2158 ATCCAGACAAAATCCCAACAGATCCAGAGATGGCTGATTAATAATTCTGAGCTCGAAGAG 2217  
Qy 2221 AAACATGGGAGGAGCCCAAGTCTCAGCCAGCACTTGAAGAGTGCACCTTGAACAGAAAGAG 2280  
Db 2218 AAACATGGGAGGAGCCCAAGTCTCAGCCAGCACTTGAAGAGTGCACCTTGAACAGAAAGAG 2277  
Qy 2281 CAGCACTATGAGAAAAGATTAAAGTGTGGACATCGATTAAGAAAAGACTTGGCTGAC 2340  
Db 2278 CAGCACTATGAGAAAAGATTAAAGTGTGGACATCGATTAAGAAAAGACTTGGCTGAC 2337  
Qy 2341 AAGGAGACATGGAGAAATGATATCGAGACACAGAGAGAGGCCCATGAGAAAGGCAAA 2400  
Db 2338 AAGGAGACATGGAGAAATGATATCGAGACACAGAGAGAGGCCCATGAGAAAGGCAAA 2397  
Qy 2401 ATTCTCAGCGAACAAGAGCGATGATCAATGCTATGATTCAGAGATCGATTCCTGGAA 2460  
Db 2398 ATTCTCAGCGAACAAGAGCGATGATCAATGCTATGATTCAGAGATCGATTCCTGGAA 2457  
Qy 2461 CAGAGATTGTGGAACTGTCTGAAGCAATTAACCTTGACGAAATAGCAGTCTTTTAAAC 2520  
Db 2458 CAGAGATTGTGGAACTGTCTGAAGCAATTAACCTTGACGAAATAGCAGTCTTTTAAAC 2517  
Qy 2521 CAAGAGAACTAGAAAGGAGCCCAAGAAAGAGATGATTTCTGAATCAGGCAACAGAAATTTTAC 2580  
Db 2518 CAAGAGAACTAGAAAGGAGCCCAAGAAAGAGATGATTTCTGAATCAGGCAACAGAAATTTTAC 2577  
Qy 2581 CTGAGAGACAGAGCTGGAGAAAGTGGAGGCCAGAAACGAAATCTGGAGAGCGCTGAG 2640  
Db 2578 CTGAGAGACAGAGCTGGAGAAAGTGGAGGCCAGAAACGAAATCTGGAGAGCGCTGAG 2637  
Qy 2641 AAGATCAGCCACCAAGACACAGTGAACAAATCGGCTGCTGAACTGAGACAAAGATTG 2700  
Db 2638 AAGATCAGCCACCAAGACACAGTGAACAAATCGGCTGCTGAACTGAGACAAAGATTG 2697  
Qy 2701 CGGAGAGTCACTTGAAGACAGAGAGCAAGAACTGAGAGCTCAAGCGCTCAACAGAG 2760  
Db 2698 CGGAGAGTCACTTGAAGACAGAGAGCAAGAACTGAGAGCTCAAGCGCTCAACAGAG 2757  
Qy 2761 CTACAGCTCTCCCTGAGAGAGCGGAGTCAAGTTGACAGCCCTGAGAGCTGACCGGCG 2820  
Db 2758 CTACAGCTCTCCCTGAGAGAGCGGAGTCAAGTTGACAGCCCTGAGAGCTGACCGGCG 2817  
Qy 2821 GCCCTGAGAGAGCACTTTCGACAGGCGAAGACAGAGCTGAGAGACACAGAGAAAGCT 2880  
Db 2818 GCCCTGAGAGAGCACTTTCGACAGGCGAAGACAGAGCTGAGAGACACAGAGAAAGCT 2877  
Qy 2881 GAAGAGAGATTCAGGCACTCAGGCAATAGAGTGAATTCAGGCGAAATTTGATGCT 2940  
Db 2878 GAAGAGAGATTCAGGCACTCAGGCAATAGAGTGAATTCAGGCGAAATTTGATGCT 2937  
Qy 2941 CTTTCGTAAGAGCTGTACTGTATCAGAGCTTGAAGAGAGCTAAACCAAGCTGAGACGAG 3000  
Db 2938 CTTTCGTAAGAGCTGTACTGTATCAGAGCTTGAAGAGAGCTTAAACCAAGCTGAGACGAG 2997  
Qy 3001 GACAAAGCTGAACTCAACAAACAAATCTTCACTTGTCCAAACAACTCGATGAGGCTTCT 3060  
Db 2998 GACAAAGCTGAACTCAACAAACAAATCTTCACTTGTCCAAACAACTCGATGAGGCTTCT 3057  
Qy 3061 GGGGCGCAACGAGAGATTGTACAACTGCGAAGTGAAGTGAACATCTCCGCGGAGATC 3120  
Db 3058 GGGGCGCAACGAGAGATTGTACAACTGCGAAGTGAAGTGAACATCTCCGCGGAGATC 3117

Qy 3121 ACGGAACGAGATGACAGCTTACCAAGCAAGAACAGATGAGGCTCTGAAAGACACAG 3180  
Db 3118 ACGGAACGAGATGACAGCTTACCAAGCAAGAACAGATGAGGCTCTGAAAGACACAG 3177  
Qy 3181 TGCAACATGCTGGAGGAAACAGGTCAATGATTTGGAGGCGCTTAACGATGAGTCTTAGAA 3240  
Db 3178 TGCAACATGCTGGAGGAAACAGGTCAATGATTTGGAGGCGCTTAACGATGAGTCTTAGAA 3237  
Qy 3241 AAAGAGCGGAGTGGAGAGCGCTGGAGAGCGCTCGGAGTGAATGAGAAATCCAGTTTGA 3300  
Db 3238 AAAGAGCGGAGTGGAGAGCGCTGGAGAGCGCTCGGAGTGAATGAGAAATCCAGTTTGA 3297  
Qy 3301 TGTGAGGTTGAGAGCTGCAAGAGATGCTGACACCGAGAAACAGAGAGCGGCGAGAGCC 3360  
Db 3298 TGTGAGGTTGAGAGCTGCAAGAGATGCTGACACCGAGAAACAGAGAGCGGCGAGAGCC 3357  
Qy 3361 GATCAGCGGATCACCGAGCTCTGCGCAGGTGTGAGTGGCAATGGAAGAGACAAAGCT 3420  
Db 3358 GATCAGCGGATCACCGAGCTCTGCGCAGGTGTGAGTGGCAATGGAAGAGACAAAGCT 3417  
Qy 3421 GAGATTCTCGCTGTCAGACAGGCTCTCAAGAGAGAAAGCTGAAGGCGGAGAGCGCTCT 3480  
Db 3418 GAGATTCTCGCTGTCAGACAGGCTCTCAAGAGAGAAAGCTGAAGGCGGAGAGCGCTCT 3477  
Qy 3481 GACAACTCAATGAATCTGGAGAGAGAGATGCTATGCTTGAATGAATGCCGAAGCTTAA 3540  
Db 3478 GACAACTCAATGAATCTGGAGAGAGAGATGCTATGCTTGAATGAATGCCGAAGCTTAA 3537  
Qy 3541 CAGCAGAGGCTGGAGACTGAAACGAGAGCTCAAAACAGAGGCTTCTGGAAGAGACCAAGCT 3600  
Db 3538 CAGCAGAGGCTGGAGACTGAAACGAGAGCTCAAAACAGAGGCTTCTGGAAGAGACCAAGCT 3597  
Qy 3601 TTAACAGAGAGATGAGACTGCAAGAAATCAACTTTCCGCTGACCTCAAGAGACTGCA 3660  
Db 3598 TTAACAGAGAGATGAGACTGCAAGAAATCAACTTTCCGCTGACCTCAAGAGACTGCA 3657  
Qy 3661 GAAGCTTAATGAGGCTGATCTTAAGAGACAGAAAGAAAGTGAAGTGAAGTATCAGCTG 3720  
Db 3658 GAAGCTTAATGAGGCTGATCTTAAGAGACAGAAAGAAAGTGAAGTGAAGTATCAGCTG 3717  
Qy 3721 GAAACATTAAGGTTCTTATTTCTATGAAAGAGTGAATTCGAAGGCACTATTTCTCAA 3780  
Db 3718 GAAACATTAAGGTTCTTATTTCTATGAAAGAGTGAATTCGAAGGCACTATTTCTCAA 3777  
Qy 3781 CAACCAAACTCAATGATTTTCTGCAAGCCAAATGAGACCACTGCTAAGAAAGAAAG 3840  
Db 3778 CAACCAAACTCAATGATTTTCTGCAAGCCAAATGAGACCACTGCTAAGAAAGAAAG 3837  
Qy 3841 GTTCTCTGAGTACAAATGAGCTGAAGCTGACCCTGAGAGAGAAAGTCCGCTGTGCA 3900  
Db 3838 GTTCTCTGAGTACAAATGAGCTGAAGCTGACCCTGAGAGAGAAAGTCCGCTGTGCA 3897  
Qy 3901 GAGCTAGAGGAACCCCTTCAAGAGACCCGATGAGCTCCGCTGAGAGAAAGCT 3960  
Db 3898 GAGCTAGAGGAACCCCTTCAAGAGACCCGATGAGCTCCGCTGAGAGAAAGCT 3957  
Qy 3961 GCCCAGCGAAGCAACGACACCAACCACTTCAACGCGCAGGCAACCGAGGCAACAG 4020  
Db 3958 GCCCAGCGAAGCAACGACACCAACCACTTCAACGCGCAGGCAACCGAGGCAACAG 4017  
Qy 4021 ATGCGCATGTCGCGCATGTCGCGTGGCAAGACACAGCCCAAGTGCATGAGCTTGTG 4080  
Db 4018 ATGCGCATGTCGCGCATGTCGCGTGGCAAGACACAGCCCAAGTGCATGAGCTTGTG 4077  
Qy 4081 GCCCGCGCATTCAGCGCGGAGAAAGAGTCTTCAATCCAGAGAAATTTAATCGGCGCTT 4140  
Db 4078 GCCCGCGCATTCAGCGCGGAGAAAGAGTCTTCAATCCAGAGAAATTTAATCGGCGCTT 4137  
Qy 4141 AAGGAACGATCACCACCAATATTTCTCAACCGATTCAGAGTGAAGTGAAGTGAAGCC 4200  
Db 4138 AAGGAACGATCACCACCAATATTTCTCAACCGATTCAGAGTGAAGTGAAGTGAAGCC 4197



XX OS Homo sapiens.  
XX Key Location/Qualifiers  
XX CDS 1..5877  
XX /tag= a  
XX /product= "Kinase"  
XX MO200259325-A2.  
XX 01-AUG-2002.  
XX 20-DEC-2001; 2001WO-US50497.  
XX 27-DEC-2000; 2000US-258335P.  
XX (LEXI-) LEXICON GENETICS INC.  
XX Yu X, Miranda W, Friiddle CJ;  
XX WPI; 2002-599796/64.  
XX P-PSDB; ABB81928.  
XX Novel polynucleotide encoding human proteins that are structurally  
XX similar to animal kinases, useful for drug screening, diagnosis, in  
XX gene therapy of disorders and diseases e.g. cancer and pharmacogenomic  
XX applications -  
XX  
XX Disclosure; Page 44-45; 50pp; English.  
XX  
XX The invention relates to a novel human protein that shares structural  
XX similarity with animal kinases, including serine-threonine kinases,  
XX particularly Cdkron rho-interacting kinases. The proteins of the  
XX invention have nootropic and cytoskeletal activity. The polynucleotides may  
XX have a use in gene therapy. The encoded novel polypeptides are useful for  
XX generating antibodies, as reagents in diagnostic assays, for identifying  
XX other cellular gene products related to NHR and as reagents in assays for  
XX screening for compounds that are useful in the treatment of mental,  
XX biological or medical disorders and diseases including cancer. The  
XX sequence encodes a novel human kinase of the invention.  
XX  
XX Sequence 5877 BP; 1680 A; 1433 C; 1583 G; 1181 T; 0 other;  
XX  
XX Query Match 94.7%; Score 5837; DB 24; Length 5877;  
XX Best Local Similarity 100.0%; Pred. No. 0;  
XX Matches 5837; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 ATGTTGAAGTCAAAATATGAGCGCGAATCTTTGATGCTGTGCTGAACCAT 60  
XX 1 ATGTTGAAGTCAAAATATGAGCGCGAATCTTTGATGCTGTGCTGAACCAT 60  
XX 61 GCCAGCGCGGCTCCAGGCTGAATCTTTCTTCAAGGGAACACCTTTATGACTCA 120  
XX 61 GCCAGCGCGGCTCCAGGCTGAATCTTTCTTCAAGGGAACACCTTTATGACTCA 120  
XX 121 CAGCAGATGCTCTCTTTCCCGAGAAGATATTAGATGCCCTCTTTGTTCTTTGAA 180  
XX 121 CAGCAGATGCTCTCTTTCCCGAGAAGATATTAGATGCCCTCTTTGTTCTTTGAA 180  
XX 181 GAATGAGTCAAGCTCTCTGATGAGATTAAGCACTGAGCAACTTTGTCGGAATAT 240  
XX 181 GAATGAGTCAAGCTCTCTGATGAGATTAAGCACTGAGCAACTTTGTCGGAATAT 240  
XX 241 TCCGACACCATAGCTGATTAACAGAGCTCAGCTTCGCAAGGCTCA 300  
XX 241 TCCGACACCATAGCTGATTAACAGAGCTCAGCTTCGCAAGGCTCA 300  
XX 301 AGCTTTAGAGTTGTGTCTCTTGTGTAAGTGAAGTGTGTAAGAGAAAGCAACG 360  
XX 301 AGCTTTAGAGTTGTGTCTCTTGTGTAAGTGAAGTGTGTAAGAGAAAGCAACG 360  
XX 361 GAGATCTAGCTAGAGATGAGTAAGAGAGCTTTATGGCCGAGAGAGGTTTCA 420  
XX 361 GAGATCTAGCTAGAGATGAGTAAGAGAGCTTTATGGCCGAGAGAGGTTTCA 420

Db 361 GACATCTATGCTATGAAAGTATGAAAGAGAGCTTTATGGCCGAGAGAGGTTTCA 420  
Qy 421 TTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Db 421 TTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Qy 481 CAGTATGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
Db 481 CAGTATGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
Qy 541 GACTTGTCTCTCTTGTGATATGATATGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
Db 541 GACTTGTCTCTCTTGTGATATGATATGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
Qy 601 TACCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Db 601 TACCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Qy 661 GACATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
Db 661 GACATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
Qy 721 GATCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
Db 721 GATCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
Qy 781 CCAGATTAACATGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
Db 781 CCAGATTAACATGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
Qy 841 GGCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 841 GGCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Qy 901 TCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 901 TCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Qy 961 TTTTGAATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 961 TTTTGAATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Qy 1021 TTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
Db 1021 TTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
Qy 1081 TCTAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
Db 1081 TCTAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
Qy 1141 TCTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
Db 1141 TCTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
Qy 1201 CCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
Db 1201 CCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
Qy 1261 TACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db 1261 TACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Qy 1321 CCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
Db 1321 CCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
Qy 1381 TCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Db 1381 TCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Qy 1441 GAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
Db 1441 GAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500

```
QY 1501 TCCTCTTGGAGCAGACCTTGTCTTCTTCAATCAACAGATGACAGTAAAGCTTAAAGCGAAGT 1560
DB 1501 TCCTCTTGGAGCAGACCTTGTCTTCTTCAATCAACAGATGACAGTAAAGCTTAAAGCGAAGT 1560
QY 1561 TTGGAGCAGCAGAGTGGAGGTGTCCAGAGAGATGACAAAGCAGTGCAGCTTCTCAT 1620
DB 1561 TTGGAGCAGCAGAGTGGAGGTGTCCAGAGAGATGACAAAGCAGTGCAGCTTCTCAT 1620
QY 1621 GATATCAGAGCAGAGCCCGAAGCTCCAGAAATCAAAGAGCAGAGTCCAGGCTCAA 1680
DB 1621 GATATCAGAGCAGAGCCCGAAGCTCCAGAAATCAAAGAGCAGAGTCCAGGCTCAA 1680
QY 1681 GTGGAGAAATGAGGTGATGATGAATCACTTGAAGAGATCTTGTCTCAGCAAGAGA 1740
DB 1681 GTGGAGAAATGAGGTGATGATGAATCACTTGAAGAGATCTTGTCTCAGCAAGAGA 1740
QY 1741 CGAGAGATCTCTACAAATCTGAGCTGAGAGATCTCGCTTGTCTGTGAAGATTCCAG 1800
DB 1741 CGAGAGATCTCTACAAATCTGAGAGATCTCGCTTGTCTGTGAAGATTCCAG 1800
QY 1801 CGGAAAGCGACAGAAATGTCAGCATAACTGTTGAAGCTTAAGATCAAGGAGCTGAA 1860
DB 1801 CGGAAAGCGACAGAAATGTCAGCATAACTGTTGAAGCTTAAGATCAAGGAGCTGAA 1860
QY 1861 GTGGAGAAATATCCGAAATCTGAGAAATCAATGCTGAGCAGAGCTCAAAATTCAGAG 1920
DB 1861 GTGGAGAAATATCCGAAATCTGAGAAATCAATGCTGAGCAGAGCTCAAAATTCAGAG 1920
QY 1921 CTCACAGAGAAATGAGAGAGCTGTAAAGCCGAGACGAGAGCCACCGAGCTGCTCAG 1980
DB 1921 CTCACAGAGAAATGAGAGAGCTGTAAAGCCGAGACGAGAGCCACCGAGCTGCTCAG 1980
QY 1981 AATATCCGCGCAAGAGAGCGAGCCGAGAGAGCTGAGAGAGCTGAGAAACCGAGAG 2040
DB 1981 AATATCCGCGCAAGAGAGCGAGCCGAGAGAGCTGAGAGAGCTGAGAAACCGAGAG 2040
QY 2041 GATTCCTTCTGAAAGGCTACGAAAGAAAGCTGTGTGAAGCTGAGAGAGCCGCCATTCCTG 2100
DB 2041 GATTCCTTCTGAAAGGCTACGAAAGAAAGCTGTGTGAAGAGCTGAGAGAGCCGCCATTCCTG 2100
QY 2101 GAGAACAAAGTAAAGAGCTAGAGACCATGAGCCATGAGAGAAACGAGCTGAAGAGTGA 2160
DB 2101 GAGAACAAAGTAAAGAGCTAGAGACCATGAGCCATGAGAGAAACGAGCTGAAGAGTGA 2160
QY 2161 ATCCAGACAAAAATCCCAACAGATCCAGAGATGCTGATTAATTCGTGAGCTCGAAGAG 2220
DB 2161 ATCCAGACAAAAATCCCAACAGATCCAGAGATGCTGATTAATTCGTGAGCTCGAAGAG 2220
QY 2221 AAAACATCCGAGAGCCCAAGCTTCAGCCGAGCACTTAAGAGTCACTGAAACAGAAAGAG 2280
DB 2221 AAAACATCCGAGAGCCCAAGCTTCAGCCGAGCACTTAAGAGTCACTGAAACAGAAAGAG 2280
QY 2281 CAGCACTATGAGGAAAGATTAAGTGTGACAAATCAGATTAAGAAAGAAACCTGCTGAC 2340
DB 2281 CAGCACTATGAGGAAAGATTAAGTGTGACAAATCAGATTAAGAAAGAAACCTGCTGAC 2340
QY 2341 AAGAGACACTGAGAAACATGATGACAGACACGAGAGAGAGCCCATGAGAAAGGCAAA 2400
DB 2341 AAGAGACACTGAGAAACATGATGACAGACACGAGAGAGAGCCCATGAGAAAGGCAAA 2400
QY 2401 ATTCTCAGAGAACAGAGGCGATGATCAATGCTATGATTCCAAGATCAGATCCCTGGA 2460
DB 2401 ATTCTCAGAGAACAGAGGCGATGATCAATGCTATGATTCCAAGATCAGATCCCTGGA 2460
QY 2461 CAGAGGATTTGGAACTGTCTGAAAGCCAAATACTTGACAAATAGCAGTCTTTTACC 2520
DB 2461 CAGAGGATTTGGAACTGTCTGAAAGCCAAATACTTGACAAATAGCAGTCTTTTACC 2520
QY 2521 CAAGAGAACTAGAGGCCCCAAGAGAGATGATTTTGAACTCAGGCAACAGAAATTTTAC 2580
DB 2521 CAAGAGAACTAGAGGCCCCAAGAGAGATGATTTTGAACTCAGGCAACAGAAATTTTAC 2580
QY 2581 CTGAGACACAGGCTGGAGAGTTTGGAGGCCAGAAACGAGAACTGAGAGCAGCTGAG 2640
DB 2581 CTGAGACACAGGCTGGAGAGTTTGGAGGCCAGAAACGAGAACTGAGAGCAGCTGAG 2640
QY 2641 AAGATCAGCCCAACAACAACAGTGAACAAGATCCGCTGTGAACTGAGACAGAATG 2700
DB 2641 AAGATCAGCCCAACAACAACAGTGAACAAGATCCGCTGTGAACTGAGACAGAATG 2700
QY 2701 CGGAGAGTCAATTTAAGACACGAGAGACGAAACTGAGACTCAAGCCGCTCACAGAG 2760
DB 2701 CGGAGAGTCAATTTAAGACACGAGAGACGAAACTGAGACTCAAGCCGCTCACAGAG 2760
QY 2761 CTACAGCTCTCCCTGAGAGAGCCGAGTCAAGTTCAGAGCTTCGAGGCTGACGAGCG 2820
DB 2761 CTACAGCTCTCCCTGAGAGAGCCGAGTCAAGTTCAGAGCTTCGAGGCTGACGAGCG 2820
QY 2821 GCCCTGAGAGCCAGCTTTCGACGAGCGAAGACAGAGCTGAGAGACCAAGCAGAGCT 2880
DB 2821 GCCCTGAGAGCCAGCTTTCGACGAGCGAAGACAGAGCTGAGAGACCAAGCAGAGCT 2880
QY 2881 GAAGAGAGATCCAGGCACTCAGCGGACATTAAGATGAATCCAGGCGCAATTTGATGCT 2940
DB 2881 GAAGAGAGATCCAGGCACTCAGCGGACATTAAGATGAATCCAGGCGCAATTTGATGCT 2940
QY 2941 CTTCTGAAGAGCTGTACTGTAAATCAGAGACTTGAAGAGAGAGCTTAAACAGTGAACGAG 3000
DB 2941 CTTCTGAAGAGCTGTACTGTAAATCAGAGACTTGAAGAGAGAGCTTAAACAGTGAACGAG 3000
QY 3001 GACACGCTGAATCTCAACAACCAAACTTCTTCAACCAACTGATGAAGCTTCT 3060
DB 3001 GACACGCTGAATCTCAACAACCAAACTTCTTCAACCAACTGATGAAGCTTCT 3060
QY 3061 GCGCGCAAGAGAGATTTGTAACAATGCGAAGTGAAGTGAACATTCGCGCGGAGATC 3120
DB 3061 GCGCGCAAGAGAGATTTGTAACAATGCGAAGTGAAGTGAACATTCGCGCGGAGATC 3120
QY 3121 ACGGAAACGAGAGATGAGCTTACAGCCAGAAACCAAGATGAGAGCTTGAAGACACAG 3180
DB 3121 ACGGAAACGAGAGATGAGCTTACAGCCAGAAACCAAGATGAGAGAGCTTGAAGACACAG 3180
QY 3181 TGCACCAATGCTGAGAGAACAGGTCAATGATTTGAGAGCCCTTAAACGATGAGCTGTAGA 3240
DB 3181 TGCACCAATGCTGAGAGAACAGGTCAATGATTTGAGAGCCCTTAAACGATGAGCTGTAGA 3240
QY 3241 AAAAGCGGCAATGAGAGGCTCTGAGAGAGGCTCTGAGAGAGAGAAATCCCAATTTGAG 3300
DB 3241 AAAAGCGGCAATGAGAGGCTCTGAGAGAGGCTCTGAGAGAGAGAAATCCCAATTTGAG 3300
QY 3301 TGTGCGGTTTCAAGAGCTGAGAGATGCTGACACCGAGAAACAGAGCAGGCGAGAGCC 3360
DB 3301 TGTGCGGTTTCAAGAGCTGAGAGATGCTGACACCGAGAAACAGAGCAGGCGAGAGCC 3360
QY 3361 GATCAGCGGATTAACGAGCTCTGCGCAGGTGTGAGCTGAGAGAGAGCAAGGCT 3420
DB 3361 GATCAGCGGATTAACGAGCTCTGCGCAGGTGTGAGCTGAGAGAGAGCAAGGCT 3420
QY 3421 GAGATTTCTGCTCTGAGAGGCTCTCAAGAGCAGAAAGCTGAAGGCGAGGCTCTCT 3480
DB 3421 GAGATTTCTGCTCTGAGAGGCTCTCAAGAGCAGAAAGCTGAAGGCGAGGCTCTCT 3480
QY 3481 GACAAAGCTCAATGACCTTGAAGAGAGATGATCTTGAATGAATGACCCGAGCTTA 3540
DB 3481 GACAAAGCTCAATGACCTTGAAGAGAGATGATCTTGAATGAATGACCCGAGCTTA 3540
QY 3541 CAGCAGAGCTGAGAGACTGAACGAGAGCTCAACAGAGGCTTCTGAAGAGCAGCCAAA 3600
DB 3541 CAGCAGAGCTGAGAGACTGAACGAGAGCTCAACAGAGGCTTCTGAAGAGCAGCCAAA 3600
QY 3601 TTACAGCAGAGATGAGCTTCAAGAAATACATTTTCGCTGAGCTCAAGAGCTGGA 3660
DB 3601 TTACAGCAGAGATGAGCTTCAAGAAATACATTTTCGCTGAGCTCAAGAGCTGGA 3660
QY 3661 GAGCTTAGATCGGCTGATCTAATGAAGCAGAAAGAGTGAAGCTTGAAGATTCAGCTG 3720
```



Db	3661	GAAGCTCTAGATGGGCTGATCTACTGAAGACAGAAAGAGTGAAGCTTGAGATCAAGCTG	3720
OY	3721	GAAGCAATTCAGGTTCTTAATTCCTCAGTAAAAAGTGAATATGAAGCACTATTCTCAA	3780
Db	3721	GAAGCAATTCAGGTTCTTAATTCCTCAGTAAAAAGTGAATATGAAGCACTATTCTCAA	3780
OY	3781	CAAAACCAACTCTATTAATTTTCTGCAAGCCAAATATGACCAACTGCTTAAAAAGAAAAG	3840
Db	3781	CAAAACCAACTCTATTAATTTTCTGCAAGCCAAATATGACCAACTGCTTAAAAAGAAAAG	3840
OY	3841	GTTCCTCTCAGTACCAATGAGCTGAAGGCGCCCTGGAGAGAGAGAAAGCTCGCTGAGCA	3900
Db	3841	GTTCCTCTCAGTACCAATGAGCTGAAGGCGCCCTGGAGAGAGAGAAAGCTCGCTGAGCA	3900
OY	3901	GAGCTTAGAGAAAGCCCTTCAGAAAGACCCGCACTCGAGCTCGGCTCGCCCGGAGAAAGCT	3960
Db	3901	GAGCTTAGAGAAAGCCCTTCAGAAAGACCCGCACTCGAGCTCGGCTCGCCCGGAGAAAGCT	3960
OY	3961	GCCCAACCGAAAGCAACGAGCCACCCCAACCCATCAGCCAGCCACCCGAGGCAAG	4020
Db	3961	GCCCAACCGAAAGCAACGAGCCACCCCAACCCATCAGCCAGCCACCCGAGGCAAG	4020
OY	4021	ATCGCCATATCGGCAATGATGCGGCTGCGCAGAGCAACGAGCCCAATGCTGAGCGTCTG	4080
Db	4021	ATCGCCATATCGGCAATGATGCGGCTGCGCAGAGCAACGAGCCCAATGCTGAGCGTCTG	4080
OY	4081	GGCCGCCCATCCAGCCGAGAAAGAGTCTTCAACTCAGAGAAATTTAGTCGGCGCTT	4140
Db	4081	GGCCGCCCATCCAGCCGAGAAAGAGTCTTCAACTCAGAGAAATTTAGTCGGCGCTT	4140
OY	4141	AAGGAACCGATGACACCAATATTTCTCAGCCGATTAACGTMAGACTGAACATGCGAGCC	4200
Db	4141	AAGGAACCGATGACACCAATATTTCTCAGCCGATTAACGTMAGACTGAACATGCGAGCC	4200
OY	4201	ACAAAGTGTGCTGTGTCTGGATTAACGCTGACTTTTGAAGGCGCAGGCACTCAATGCTC	4260
Db	4201	ACAAAGTGTGCTGTGTCTGGATTAACGCTGACTTTTGAAGGCGCAGGCACTCAATGCTC	4260
OY	4261	GAATGTCAAGTATGTGTCAACCCAGTGTCCACGTCGTTGCCAGCACTCGCGGCTTG	4320
Db	4261	GAATGTCAAGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4320
OY	4321	CCTGCTGAATTTGCCACACACTTCAACGAGGCTTTCGCGGTGACAAATGAATCCCA	4380
Db	4321	CCTGCTGAATTTGCCACACACTTCAACGAGGCTTTCGCGGTGACAAATGAATCCCA	4380
OY	4381	GGTCTCCGAGCAACGAGGAGCCGAGAGGAGTGTGACCTGGAAGGCTGTGATGAAGGTCGCC	4440
Db	4381	GGTCTCCGAGCAACGAGGAGCCGAGAGGAGTGTGACCTGGAAGGCTGTGATGAAGGTCGCC	4440
OY	4441	AGGAATTAACAAACGAGGACAGCAAGGCTGGGACAGAAATGACATTTGCTCGAGGGATCA	4500
Db	4441	AGGAATTAACAAACGAGGACAGCAAGGCTGGGACAGAAATGACATTTGCTCGAGGGATCA	4500
OY	4501	AAAGTCTCATTTATGACAAATGAAGCCAGAGAAAGCTGACAGAGGCGGAGGAAGATTT	4560
Db	4501	AAAGTCTCATTTATGACAAATGAAGCCAGAGAAAGCTGACAGAGGCGGAGGAAGATTT	4560
OY	4561	GAGCTGTGCGCTCCCGAGGGGAGTATCTAATTCATGTGTGCGCTTCCGATCTC	4620
Db	4561	GAGCTGTGCGCTCCCGAGGGGAGTATCTAATTCATGTGTGCGCTTCCGATCTC	4620
OY	4621	GCAAAATACAGCAAAAGCAGATGTCCCATATCTGAATGTGAATCTCAACCCGACAC	4680
Db	4621	GCAAAATACAGCAAAAGCAGATGTCCCATATCTGAATGTGAATCTCAACCCGACAC	4680
OY	4681	ACCTGCTGTGCGCGGAGAACCTCTCATCTTGTAGCTCCAGGCTCCCTGACAAACGCGC	4740
Db	4681	ACCTGCTGTGCGCGGAGAACCTCTCATCTTGTAGCTCCAGGCTCCCTGACAAACGCGC	4740
OY	4741	TGGGTACCGCGCTTAGATCAGTTGTGCAAGGTGGAGGTTTCTTAGGAAACACAGAA	4800
Db	4741	TGGGTACCGCGCTTAGATCAGTTGTGCAAGGTGGAGGTTTCTTAGGAAACACAGAA	4800

[illegible]



RESULT 10  
AAD39191

ID AAD39191 standard; cDNA; 6574 BP.

AC AAD39191;

DT 04-OCT-2002 (first entry)

XX Human MDPK cDNA.

XX Human; myotonic dystrophy type protein kinase; MDPK; 13245 protein;  
KW tumorigenesis; tumour growth; tumour metastasis; viral infection;  
KW skeletal muscle disorder; muscular dystrophy; myotonic dystrophy;  
KW immune disorder; neoplastic disorder; gene therapy; gene; ss.

OS Homo sapiens.

XX Location/Qualifiers

FT 5'UTR 1..18

FT CDS /tag= a

FT /tag= b

FT /product= "Human MDPK protein"

FT /tag= c

FT /note= "This region is specifically referred

FT in claim 1 as SEQ ID NO:3"

FT 3'UTR 6181..6574

FT /tag= d

PN WO020234896-A2.

PD 02-MAY-2002.

PF 23-OCT-2001; 2001WO-US50636.

PR 23-OCT-2000; 2000US-242429P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Kapeller-Libermann R;

DR WPI; 2002-479720/51.

DR P-Psdb; AAE24079.

PT Human myotonic dystrophy type protein kinase polypeptide and

PT inhibiting tumorigenesis, tumor growth, tumor metastasis and viral

PT infection

XX Claim 1; Fig 1; 148pp; English.

XX The invention relates to human myotonic dystrophy type protein kinase

XX (MDPK) polypeptides designated as 13245 and nucleic acid molecules

XX encoding such polypeptides. 13245 molecules are used to develop

XX preventing, inhibiting, alleviating or curing MDPK-related disorders.

XX Polypeptides of the invention are used to develop diagnostic and

XX therapeutic agents for 13245-mediated or related disorders such as

XX tumorigenesis, tumour growth, tumour metastasis, viral infection

XX of a cell, skeletal muscle disorders (e.g. muscular and myotonic

XX dystrophies), immune disorders and neoplastic disorders. The

XX invention is also used in gene therapy. The present sequence is

XX human MDPK cDNA.

XX Sequence 6574 BP; 1877 A; 1611 C; 1776 G; 1310 T; 0 other;

XX Query Match 91.8%; Score 5661.4; DB 24; Length 6574;

XX Best Local Similarity 95.8%; Pred. No. 0;

XX Matches 5944; Conservative 0; Mismatches 16; Indels 243; Gaps 3;

QY 1 ATGTTGAAGTTCATAATATGAGCGCGAATCCCTTGGATGCTGCTGTAACCATTT 60  
DB 19 ATGTTGAAGTTCATAATATGAGCGCGAATCCCTTGGATGCTGCTGTAACCATTT 78  
QY 61 GCCAGCCGGGCTCCAGGCTGAATCTGTTCTCCAGGAGAAACACCTTTATGACTCAA 120  
DB 79 GCCAGCCGGGCTCCAGGCTGAATCTGTTCTCCAGGAGAAACACCTTTATGACTCAA 138  
QY 121 CAGCAGATGCTCTCTTCCCGAGAGAGGATTTAGATGCTCTCTTGTCTCTTTGAA 180  
DB 139 CAGCAGATGCTCTCTTCCCGAGAGAGGATTTAGATGCTCTCTTGTCTCTTTGAA 198  
QY 181 GAATGAGTCAAGCTCTCTCTGATGAAGATTAAAGCAGTGAACATTTGTCGGAAGTAT 240  
DB 199 GAATGAGTCAAGCTCTCTGATGAAGATTAAAGCAGTGAACATTTGTCGGAAGTAT 258  
QY 241 TCCGACACCAATGCTGAGTTACAGAGCTCCAGCTTCGCAAGAGACTTCGAGTCA 300  
DB 259 TCCGACACCAATGCTGAGTTACAGAGCTCCAGCTTCGCAAGAGACTTCGAGTCA 318  
QY 301 AGCTTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 360  
DB 319 AGCTTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 378  
QY 361 GACATCTATGCTATGAAAAGTGAAGAAAGGCTTTATTTGGCCAGAGAGGTTTCA 420  
DB 379 GACATCTATGCTATGAAAAGTGAAGAAAGGCTTTATTTGGCCAGAGAGGTTTCA 438  
QY 421 TTTTGGAGAAAGCGGACATATTATCTCGAAGCACAAGCCGTGATCCCCCAATT 480  
DB 439 TTTTGGAGAAAGCGGACATATTATCTCGAAGCACAAGCCGTGATCCCCCAATT 498  
QY 481 CAGTATGCTTTCAGAGCAAAAATGACCTTTATCTGTTGATGAGATTTAGGCTGAGG 540  
DB 499 CAGTATGCTTTCAGAGCAAAAATGACCTTTATCTGTTGATGAGATTTAGGCTGAGG 558  
QY 541 GACTTGTCTCACTTTTGAATGATGAGACCAAGTTAGATGAAGAACTGATACGTTT 600  
DB 559 GACTTGTCTCACTTTTGAATGATGAGACCAAGTTAGATGAAGAACTGATACGTTT 618  
QY 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCTTCAATCTGATGAGATACGTCGA 660  
DB 619 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGCTTCAATCTGATGAGATACGTCGA 678  
QY 661 GACATCAAGCTGAGAACATTCGTGACCGGACAGGACATCAACCTGCTGATTTT 720  
DB 679 GACATCAAGCTGAGAACATTCGTGACCGGACAGGACATCAACCTGCTGATTTT 738  
QY 721 GGAATGCTGCGGAAATGAAATTCAAACAGATGATGCAAACTCCGATTGGAGAC 780  
DB 739 GGAATGCTGCGGAAATGAAATTCAAACAGATGATGCAAACTCCGATTGGAGAC 798  
QY 781 CCAGATTAAGATGCTGCTGAGATGCTGATGTAACGGGAGATGAGAAAGCACTAC 840  
DB 799 CCAGATTAAGATGCTGCTGAGATGCTGATGTAACGGGAGATGAGAAAGCACTAC 858  
QY 841 GGCTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 900  
DB 859 GGCTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 918  
QY 901 TCCCTCTTCCAGAGGAACTCTGCAAGAACTTCAATATACATTAATGAAATTTCCAG 960  
DB 919 TCCCTCTTCCAGAGGAACTCTGCAAGAACTTCAATATACATTAATGAAATTTCCAG 978  
QY 961 TTTTGAATTTCCAGATGAGCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1020  
DB 979 TTTTGAATTTCCAGATGAGCCCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1038  
QY 1021 TTGTTGTGCGGCGAGAAAGAGAGATGAAGTTTGAAGTCTTCTGCGCATCTTTCTTC 1080  
DB 1039 TTGTTGTGCGGCGAGAAAGAGAGATGAAGTTTGAAGTCTTCTGCGCATCTTTCTTC 1098  
QY 1081 TCTAAATTTGATCGAAGCAATTCGTAACCTCTCTCTCTCTCTCTCTCTCTCTCTCA 1140

Db 1099 TCTAATTTGACGTGAAACAACATTCGTACTCTCTCCCTCCCTCGTTCCACCTCTCAG 1158  
Qy 1141 TCTGACGATGACCTCTCCAAATTTTGTATGATACCAAGAAAGATTTGTGGGTTTCACTCTCT 1200  
Db 1159 TCTGACGATGACCTCTCCAAATTTTGTATGATACCAAGAAAGATTTGTGGGTTTCACTCTCT 1218  
Qy 1201 CCGTCCAGCTGAGCCCTCCAGGCTTCTCGGATGAAAGATCGCCGTTTGTGGGTTTTCG 1260  
Db 1219 CCGTCCAGCTGAGCCCTCCAGGCTTCTCGGATGAAAGATCGCCGTTTGTGGGTTTTCG 1278  
Qy 1261 TACAGCAAGGACATGGGGATTTCTGTGATGATCTGATGTTGTGTGGGTTCTGACTCC 1320  
Db 1279 TACAGCAAGGACATGGGGATTTCTGTGATGATCTGATGTTGTGTGGGTTCTGACTCC 1338  
Qy 1321 CCTGCAAGACTAGTCTCATGAAAAGAACTTCTCATCAAAACCAAGAGCTACAGAC 1380  
Db 1339 CCTGCAAGACTAGTCTCATGAAAAGAACTTCTCATCAAAACCAAGAGCTACAGAC 1398  
Qy 1381 TCTCAGGACAAGTGTCAACAGATGGAGGAGAAATGACCCGCTTACATCGAGAGTGTCA 1440  
Db 1399 TCTCAGGACAAGTGTCAACAGATGGAGGAGAAATGACCCGCTTACATCGAGAGTGTCA 1458  
Qy 1441 GAGGTGAGGCTGTGCTTATGACAGAGAGGTGAGAGGCTGAGAGCTGAGACTACAGAGA 1500  
Db 1459 GAGGTGAGGCTGTGCTTATGACAGAGAGGTGAGAGGCTGAGAGCTGAGACTACAGAGA 1518  
Qy 1501 TCCCTCTGAGCAGGACCTTGTCTACATCAAGATGCACTAGCTTAAAGCGAAGT 1560  
Db 1519 TCCCTCTGAGCAGGACCTTGTCTACATCAAGATGCACTAGCTTAAAGCGAAGT 1578  
Qy 1561 TTGAGCAGACGACCGATGAGAGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCAT 1620  
Db 1579 TTGAGCAGACGACCGATGAGAGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCAT 1638  
Qy 1621 GATATCAGAGACGAGCCGGAAGCTCCAAAGAAATCAAGAGCAGAGATACAGAGCTCAA 1680  
Db 1639 GATATCAGAGACGAGCCGGAAGCTCCAAAGAAATCAAGAGCAGAGATACAGAGCTCAA 1698  
Qy 1661 GTGAGAAATAGAGTTGATGATGATCAATGTTGAAAGAGATCTTGTCTCAGCAAGAGA 1740  
Db 1699 GTGAGAAATAGAGTTGATGATGATCAATGTTGAAAGAGATCTTGTCTCAGCAAGAGA 1758  
Qy 1741 CGGAGTGAATCTCTCGAATCTGAGCTGAGAGATGCTGGGCTTGTGTGAAAGATTCAG 1800  
Db 1759 CGGAGTGAATCTCTCGAATCTGAGCTGAGAGATGCTGGGCTTGTGTGAAAGATTCAG 1818  
Qy 1801 CGGAAAGCAGAGATGTCAGCATTAATCTGTTGAAGCTTAAGATCAAGGGAAAGCTGAA 1860  
Db 1819 CGGAAAGCAGAGATGTCAGCATTAATCTGTTGAAGCTTAAGATCAAGGGAAAGCTGAA 1878  
Qy 1861 GTGAGAGAAATGCGAAATCTGAGAAATCAATGCTGAGCAGAGCTCAAAATTCAGAG 1920  
Db 1879 GTGAGAGAAATGCGAAATCTGAGAAATCAATGCTGAGCAGAGCTCAAAATTCAGAG 1938  
Qy 1921 CTCCAAGAGAACTGAGAAAGCTGTAAAGCCAGCAGCAGAGGCCAGCTGCTGAG 1980  
Db 1939 CTCCAAGAGAACTGAGAAAGCTGTAAAGCCAGCAGCAGAGGCCAGCTGCTGAG 1980  
Qy 1981 AATATCCGCCAGGCAAGAGAGCGAGCCGAGAGGAGGCTGAGAAAGCTGCAAGACCGAG 2040  
Db 1993 AATATCCGCCAGGCAAGAGAGCGAGCCGAGAGGAGGCTGAGAAAGCTGCAAGACCGAG 2010  
Qy 2041 GATTCCTTGAAGGATCAGAAAGAGAGTGTGAGAGCTGAGAAAGCTGCAAGACCGAG 2100  
Db 2011 GATTCCTTGAAGGATCAGAAAGAGAGTGTGAGAGCTGAGAAAGCTGCAAGACCGAG 2070  
Qy 2101 GAGAAACAAGGTAAAGAGCTAGAGACCATGAGAGCTGAGAGAAACAAGATGAGATGAC 2160  
Db 2071 GAGAAACAAGGTAAAGAGCTAGAGACCATGAGAGCTGAGAGAAACAAGATGAGATGAC 2130  
Qy 2161 ATCCAGACAAATCCCAACAGATCCAGAGATGCTGATTAATTTCTGAGGCTCGAAGAG 2220  
Db 2131 ATCCAGACAAATCCCAACAGATCCAGAGATGCTGATTAATTTCTGAGGCTCGAAGAG 2190  
Qy 2221 AAACATCGGAGAGCCCAAGTCTCAGCCCAAGACCTTGAAGTGTACCTGAAACAGAAAGAG 2280  
Db 2191 AAACATCGGAGAGCCCAAGTCTCAGCCCAAGACCTTGAAGTGTACCTGAAACAGAAAGAG 2250  
Qy 2281 CAGCACTATGAGAAAGATTAAGTGTGACAAATGAGATTAAGAAAGACTGACTGAC 2340  
Db 2251 CAGCACTATGAGAAAGATTAAGTGTGACAAATGAGATTAAGAAAGACTGACTGAC 2310  
Qy 2341 AAGAGACACTGAGAGAACTGATGACAGACACGAGAGAGGCCCATGAGAGAGGCAAA 2400  
Db 2311 AAGAGACACTGAGAGAACTGATGACAGACACGAGAGAGGCCCATGAGAGAGGCAAA 2370  
Qy 2401 ATTCTAGGAAACAGAGGCGATGATCAATGCTATGAGATTCAGAGTCAATCTCTGAA 2460  
Db 2371 ATTCTAGGAAACAGAGGCGATGATCAATGCTATGAGATTCAGAGTCAATCTCTGAA 2430  
Qy 2461 CAGAGATTTGAGAACTGTCTGAGCAATAAATCTGACGAAATAGCAGTCTTTTACC 2520  
Db 2431 CAGAGATTTGAGAACTGTCTGAGCAATAAATCTGACGAAATAGCAGTCTTTTACC 2490  
Qy 2521 CAAAGGAAATGAGAGCCCAAGAGAGATTTTCTGAACTGAGGCAACAGAAATTTAC 2580  
Db 2491 CAAAGGAAATGAGAGCCCAAGAGAGATTTTCTGAACTGAGGCAACAGAAATTTAC 2550  
Qy 2581 CTGAGACACAGAGCTTGGAGGTTGAGAGCCCAAGAACCCGAAACTGAGAGAGCAGCTGAG 2640  
Db 2551 CTGAGACACAGAGCTTGGAGGTTGAGAGCCCAAGAACCCGAAACTGAGAGAGCAGCTGAG 2610  
Qy 2641 AAGATGAGCCACCAAGACCAAGTGAACAAAGATCGGCTGAGAACTGAGACCAAGATTG 2700  
Db 2611 AAGATGAGCCACCAAGACCAAGTGAACAAAGATCGGCTGAGAACTGAGACCAAGATTG 2670  
Qy 2701 CGGAGAGTCACTGAGACAGAGAGGAGGAGAACTGAGCTCAAGGCGCAGCTCAAGAG 2760  
Db 2671 CGGAGAGTCACTGAGACAGAGAGGAGGAGAACTGAGCTCAAGGCGCAGCTCAAGAG 2730  
Qy 2761 CTACAGCTCTCCCTGAGAGGCGGAGTCAAGTTGAAGCCCTGCAAGGCTCAAGGCGG 2820  
Db 2731 CTACAGCTCTCCCTGAGAGGCGGAGTCAAGTTGAAGCCCTGCAAGGCTCAAGGCGG 2790  
Qy 2821 GCCCTGAGAGAGCAGCTTCCGAGGCGGAGGAGCAGAGCTGAGAAAGACCAAGCAAGCT 2880  
Db 2791 GCCCTGAGAGAGCAGCTTCCGAGGCGGAGGAGGAGCAGAGCTGAGAAAGACCAAGCAAGCT 2850  
Qy 2881 GAAAGAGAGATCAGGACCTCAGGCAATAGAGATGAATTCAGCGCAAAATTTGATCT 2940  
Db 2851 GAAAGAGAGATCAGGACCTCAGGCAATAGAGATGAATTCAGCGCAAAATTTGATCT 2910  
Qy 2941 CTTCTGTAACGCTGTAATCTGTATCAAGACTGAGAGAGCAGCTTAAACAGCTGACCGAG 3000  
Db 2911 CTTCTGTAACGCTGTAATCTGTATCAAGACTGAGAGAGCAGCTTAAACAGCTGACCGAG 2970  
Qy 3001 GACAAAGCTGAACTCAACCAAACTTCTGCTCAAAACAATCGATGAGAGCTTCT 3060  
Db 2971 GACAAAGCTGAACTCAACCAAACTTCTGCTCAAAACAATCGATGAGAGCTTCT 3030  
Qy 3061 GGGGCCAAGCAGAGATTTGTAACAATCGGAAAGTGAATGATTCCTCCCGGAGATC 3120  
Db 3031 GGGGCCAAGCAGAGATTTGTAACAATCGGAAAGTGAATGATTCCTCCCGGAGATC 3090  
Qy 3121 ACGGAAACGAGATGATGATTCAGCCAGAGAAAGAAAGATGAGAGCTTGAAGACCAAG 3180  
Db 3091 ACGGAAACGAGATGATGATTCAGCCAGAGAAAGAAAGATGAGAGCTTGAAGACCAAG 3150  
Qy 3181 TGCACCATCTGAGAGAAACAGTCAATTTGAGAGGCTTAAACGATGAGCTGCTAGAA 3240  
Db 3151 TGCACCATCTGAGAGAAACAGTCAATTTGAGAGGCTTAAACGATGAGCTGCTAGAA 3210  
Qy 3241 AAAAGCGGAGATGAGAGCCCTGAGAGAGGCGCTGAGAGATGAATTCAGAGTTTAC 3300  
Db 3211 AAAAGCGGAGATGAGAGCCCTGAGAGAGGCGCTGAGAGATGAATTCAGAGTTTAC 3270

QY 3301 TGTGGGTTGAGAGCTGCGAGAGATGCTGGAACCCGAGAAACAGAGCAGGGCGAGAGCC 3360  
DB TGTGGGTTGAGAGCTGCGAGAGATGCTGGAACCCGAGAAACAGAGCAGGGCGAGAGCC 3330  
QY 3361 GATCAGCGGATACCCGAGATCTCGCAGAGTGTGAGAGTGGCAGTGAAGAGCAAGAGCT 3420  
DB GATCAGCGGATACCCGAGATCTCGCAGAGTGTGAGAGTGGCAGTGAAGAGCAAGAGCT 3390  
QY 3421 GAGATTCCTGCTGCGACAGAGCTCTCAAGAGAGAGAGTGAAGGCGGAGAGCTCTCT 3480  
DB GAGATTCCTGCTGCGACAGAGCTCTCAAGAGAGAGAGTGAAGGCGGAGAGCTCTCT 3450  
QY 3481 GACAGCTCAATGACCTGAGAGAGAGATGCTATGCTTGAATGAATGCTGCGAGAGCTTA 3540  
DB GACAGCTCAATGACCTGAGAGAGAGATGCTATGCTTGAATGAATGCTGCGAGAGCTTA 3510  
QY 3541 CAGCAGAGCTGAGAGCTGAGAGAGCTCAAGAGAGCTCTGAGAGAGAGAGAGAGAGCTTA 3600  
DB CAGCAGAGAGCTGAGAGCTGAGAGAGCTCAAGAGAGCTCTGAGAGAGAGAGAGAGCTTA 3570  
QY 3601 TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
DB TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3630  
QY 3661 GAGAGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720  
DB GAGAGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3690  
QY 3721 GAAAGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780  
DB GAAAGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3750  
QY 3781 CAAAGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3838  
DB CAAAGCTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3810  
QY 3839 -----AGGTTCTGTCAGTAC 3855  
DB GGTATTTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3870  
QY 3856 AATGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3915  
DB AATGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3930  
QY 3916 CTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3975  
DB CTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3990  
QY 3976 ACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4035  
DB ACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4050  
QY 4036 ATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4095  
DB ATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4110  
QY 4096 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4155  
DB GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4170  
QY 4156 CACAAATATCTCAACGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4215  
DB CACAAATATCTCAACGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4230  
QY 4216 TGTCTGATACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4275  
DB TGTCTGATACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4290  
QY 4276 TGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4335  
DB TGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4350  
QY 4291 TGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4350

QY 4336 AACACCTTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4395  
DB AACACCTTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4410  
QY 4396 GAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4455  
DB GAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4470  
QY 4456 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4515  
DB GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4530  
QY 4471 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4530  
DB GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4530  
QY 4516 GACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4575  
DB GACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4590  
QY 4531 GACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4590  
DB GACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4590  
QY 4576 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4635  
DB GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4650  
QY 4591 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4650  
DB GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4650  
QY 4636 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4695  
DB GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4695  
QY 4651 GCA----- 4653  
DB GCA----- 4653  
QY 4696 AGAAGCTTCACTTCTAGCTCCAGCTTCCCTGACAAACAGAGAGAGAGAGAGAGAG 4755  
DB AGAAGCTTCACTTCTAGCTCCAGCTTCCCTGACAAACAGAGAGAGAGAGAGAGAGAGAG 4755  
QY 4756 GAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4815  
DB GAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4815  
QY 4816 CTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4875  
DB CTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4875  
QY 4876 CTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4935  
DB CTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4935  
QY 4936 TTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4995  
DB TTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4995  
QY 4996 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5055  
DB AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5055  
QY 5056 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5115  
DB GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5115  
QY 5116 CCCAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5175  
DB CCCAATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5175  
QY 5176 GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5235  
DB GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5235  
QY 5236 CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5295  
DB CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5295  
QY 5296 TTTACCAATTTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5355  
DB TTTACCAATTTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5355  
QY 5356 TTTACCAATTTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5415  
DB TTTACCAATTTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5415  
QY 5416 GAGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5475  
DB GAGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5475

```

Db      5281  GCTCTTCCAAAGAGTTCCTGCTCTCAATCTGTGAGGGAACAGCGAGGGCAGCGAGAG 5340
Qy      5476  GAGTACTTGTGTGTTTCCAGAAATTGAGAGTTTCGAGATTCCTTTCGGAAGAGCGTAGC 5535
Db      5341  GAGTACTTGTGTGTTTCCAGAAATTGAGAGTTTCGAGATTCCTTTCGGAAGAGCGTAGC 5400
Qy      5536  CGCAGAGAGATCTTCAATGAGAGTGTGCTTACCTTTGAGCTTTGAGCTTACAGAGAGCTTAT 5595
Db      5401  CGCAGAGAGATCTTCAATGAGAGTGTGCTTACCTTTGAGCTTTGAGCTTACAGAGAGCTTAT 5460
Qy      5596  CTGTTTGTGAGCCCACTTCACTCACTGAGAGATTTAGATTCGAGGACCGCTCTCTAGCA 5655
Db      5461  CTGTTTGTGAGCCCACTTCACTCACTGAGAGATTTAGATTCGAGGACCGCTCTCTAGCA 5520
Qy      5656  GGGAGCCCTGCGCGAGCTACCTGAGACATCCCGAAGCCGCGCTACTGAGGCGCTGACATT 5715
Db      5521  GGGAGCCCTGCGCGAGCTACCTGAGACATCCCGAAGCCGCGCTACTGAGGCGCTGACATT 5580
Qy      5716  TCCTCAGAGAGATTTACTTGTGCTCTCATACCAAGATTAATTAGGGTCAATTGCTGC 5775
Db      5581  TCCTCAGAGAGATTTACTTGTGCTCTCATACCAAGATTAATTAGGGTCAATTGCTGC 5640
Qy      5776  AAGGGAAACCTCGTGAAGAGATTCGAGCACTGAGACCAAGCGGGGCGCTCCAGCTCCGCG 5835
Db      5641  AAGGGAAACCTCGTGAAGAGATTCGAGCACTGAGACCAAGCGGGGCGCTCCAGCTCCGCG 5700
Qy      5836  AGCAGCCCAACAAGAGAGGCGCAACCGATCAACAGAGACATCACCAAGCGGCTGCGC 5895
Db      5701  AGCAGCCCAACAAGAGAGGCGCAACCGATCAACAGAGACATCACCAAGCGGCTGCGC 5760
Qy      5896  TCCAGCCCAAGCGCGCGCGAGAGGCGCGCAAGCGGGAGCGCAACCAAGCGGCTGCGC 5955
Db      5761  TCCAGCCCAAGCGCGCGCGAGAGGCGCGCAAGCGGGAGCGCAACCAAGCGGCTGCGC 5820
Qy      5956  TACCGCGAGGCGCGAGCGAGCTGCGAGAGCAAGTCTCTGCGCGCGCGCTGAGAGCGA 6015
Db      5821  TACCGCGAGGCGCGAGCGAGCTGCGAGAGCAAGTCTCTGCGCGCGCGCTGAGAGCGA 5880
Qy      6016  GAGAGTGTCTCCCGCGGATCTCAAGCAAGCGGAGAGCGGCTCTCCCGGAGGCTGTTT 6075
Db      5881  GAGAGTGTCTCCCGCGGATCTCAAGCAAGCGGAGAGCGGCTCTCCCGGAGGCTGTTT 5940
Qy      6076  GAAAGCAGCAGCAGGAGGCGGCTGCTGCGGAGCGGCTGAGAGACCGCGCTGCTCCAGGTG 6135
Db      5941  GAAAGCAGCAGCAGGAGGCGGCTGCTGCGGAGCGGCTGAGAGACCGCGCTGCTCCAGGTG 6000
Qy      6136  AACAAGTCTGAGAGCAAGTCTTTC 6158
Db      6001  AACCAAGGAGAGAGGCGAGAGTGC 6023

```

## RESULT 11

AAC77568 standard; cDNA; 6609 BP.

AAC77568;

08-FEB-2001 (first entry)

Human ORFX ORF3123 polynucleotide sequence SEQ ID NO:6245.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnerability; antiparkinsonian; neurotropic; neuroprotective; anticonvulsant; osteoplastic; antileukemic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antineoplastic; antithyroid; antidiabetic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

```

KW      allery; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW      bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW      thrombosis; contraceptive; ss.
OS      Homo sapiens.
PN      MO200058473-A2.
PD      05-OCT-2000.
PF      31-MAR-2000; 2000MO-US08621.
PR      31-MAR-1999; 99US-0127607.
PR      02-APR-1999; 99US-0127636.
PR      05-APR-1999; 99US-0127728.
PR      30-MAR-2000; 2000US-0540763.
PX      (CURA-) CURAGEN CORP.
PI      Shinkets RA, Leach M,
PI      WPI; 2000-602362/57.
DR      P-PSDB; AAB43359.
XX      Novel nucleic acids and peptides derived from open reading frame X,
PT      useful for treating e.g. cancers, proliferative disorders,
PT      neurodegenerative disorders and cardiovascular disease -
PS      Claim 5; Page 5429-5433; 550pp; English.
CC      AAC7446 to AAC7760 encode the proteins given in AAB4023 to AAB43397,
CC      which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC      sequences have activities such as: cytostatic; hepatotropic; vulnerability;
CC      antiparkinsonian; neurotropic; neuroprotective;
CC      osteoplastic; anticonvulsant; antileukemic; immunosuppressant;
CC      immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC      antidiabetic; hypotensive; dermatological; immunosuppressive;
CC      antiinflammatory; antibacterial; antiviral; antifungal; antineoplastic;
CC      antithyroid; and antidiabetic. The sequences can be used for determining
CC      the presence of or predisposition to, or preventing or treating
CC      pathological conditions associated with an ORFX-associated disorder. The
CC      nucleic acids can be used to express ORFX proteins in gene therapy
CC      vectors. The proteins and nucleic acids may be used to treat cancers,
CC      proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC      graft vs host disease, cardiovascular disease, diabetes mellitus,
CC      hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC      erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC      bacterial or fungal infection, immunodeficiency (SCID), AIDS, viral,
CC      allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC      nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC      coagulation; to inhibit thrombosis; and as a contraceptive.
XX      Sequence 6609 BP; 1694 A; 1779 C; 1646 G; 1490 T; 0 other;
SQ

```

Query Match 62.5%; Score 3855; DB 21; Length 6609;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 3858; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

Qy      2303  AAGTGTGAGCAATAGATTAAGAAAGCTGCTGAGCAAGAGACATCGAAGACATCA 2362
Db      299  AGGTGTGAGCAATAGATTAAGAAAGCTGCTGAGCAAGAGACATCGGAGAACATCA 358
Qy      2363  TCGAGAGACAGAGAGAGGAGCGCCATGAGAGAGGCAAAATTTCTAGCGAAGAGAGCGA 2422
Db      359  TCGAGAGACAGAGAGAGGAGCGCCATGAGAGAGGCAAAATTTCTAGCGAAGAGAGCGA 418
Qy      2423  TATCAATCTATGATTCAGATTCAGATCCCTGGAACAGAGATTGGAACGTCTG 2482
Db      419  TATCAATCTATGATTCAGATTCAGATCCCTGGAACAGAGATTGGAACGTCTG 478
Qy      2483  AAGCCATTAACCTTGAGCAAAATAGACAGCTTTTAAACCAAGAAAGAAAGAGCCCAAG 2542
Db      479  AAGCCATTAACCTTGAGCAAAATAGACAGCTTTTAAACCAAGAAAGAAAGAGCCCAAG 538

```

QY 2543 AAGAGATGATTTCTGAATCTGAGGCAAGAAAATTTTACTCTTGAGACACAGGCTGGAACT 2602  
Db 539 AAGAGATGATTTCTGAATCTGAGGCAAGAAAATTTTACTCTTGAGACACAGGCTGGAACT 598  
QY 2603 TGGAGGCCCAAGAACCGAAAATCTGAGAGCAGCTGAGAAATGATGACCAAGAACACA 2662  
Db 599 TGGAGGCCCAAGAACCGAAAATCTGAGAGCAGCTGAGAAATGATGACCAAGAACACA 658  
QY 2663 GTGACAAAGAAATCGGCTGCTGGAATCTGAGACAAAGATTCGGGAGGTCTAGCTAGAGCAGC 2722  
Db 659 GTGACAAAGAAATCGGCTGCTGGAATCTGAGACAAAGATTCGGGAGGTCTAGCTAGAGCAGC 718  
QY 2723 AGGAGCAGAACTGAGAGCTTGAAGGCGCAAGCTCACAGCTACAGCTCTCTCTGAGAGC 2782  
Db 719 AGGAGCAGAACTGAGAGCTTGAAGGCGCAAGCTCACAGCTACAGCTCTCTCTGAGAGC 778  
QY 2783 GCGAGTCAAGTTCAGAGCCTTGCAGGCTGCAAGGAGGCGCTTGGAGAGCAGCTTCGCC 2842  
Db 779 GCGAGTCAAGTTCAGAGCCTTGCAGGCTGCAAGGAGGCGCTTGGAGAGCAGCTTCGCC 838  
QY 2843 AGGCGAAGACAGAGCTTGAAGAGACACACAGAGAGCTGAAGAGATTCAGAGCACTCA 2902  
Db 839 AGGCGAAGACAGAGCTTGAAGAGAGACACACAGAGAGCTGAAGAGATTCAGAGCACTCA 898  
QY 2903 CGGACATAGAGATGAATTCAGCGCAAAATTTGATGCTCTTCTGTAACAGCTTACTGTAA 2962  
Db 899 CGGACATAGAGATGAATTCAGCGCAAAATTTGATGCTCTTCTGTAACAGCTTACTGTAA 958  
QY 2963 TCACAGACCTGAGAGAGCAGTAAACAGCTGACCGAGACAAAGCTGAACCTCAACACC 3022  
Db 959 TCACAGACCTGAGAGAGCAGTAAACAGCTGACCGAGAGCAACAGCTGAACCTCAACACC 1018  
QY 3023 AAAACCTTCACTTGTCCAAACAACCTGATGAGGCTTCTGCGCAACGACAGATTTGTAC 3082  
Db 1019 AAAACCTTCACTTGTCCAAACAACCTGATGAGGCTTCTGCGCAACGACAGATTTGTAC 1078  
QY 3083 AACTGGAAGTAAAGTGAACCAATCTCGCGCGGAGATCAAGAGAGAGATTCAGAGCTTA 3142  
Db 1079 AACTGGAAGTAAAGTGAACCAATCTCGCGCGGAGATCAAGAGAGAGATTCAGAGCTTA 1138  
QY 3143 CCAGCCAGAAAGCAACGATGAGAGCTCTGAAGACCAAGTGAACCAAGTCTGAGAGAACAG 3202  
Db 1139 CCAGCCAGAAAGCAACGATGAGAGCTCTGAAGACCAAGTGAACCAAGTCTGAGAGAACAG 1198  
QY 3203 TCATGGAATTTGAGAGGCCCTAAACGATGAGCTCTGAGAAAAGAGGCGCAGTGGAGGCT 3262  
Db 1199 TCATGGAATTTGAGAGGCCCTAAACGATGAGCTCTGAGAAAAGAGGCGCAGTGGAGGCT 1258  
QY 3263 GGAGGAGCGTCTCGGCTGATGAGAAATCCAGTTTGAAGTGTGGGTTGAGAGCTGACAG 3322  
Db 1259 GGAGGAGCGTCTCGGCTGATGAGAAATCCAGTTTGAAGTGTGGGTTGAGAGCTGACAG 1318  
QY 3323 GATCTCTGACACCGAGAAACAGAGCAGAGGCGAGACCGATCAGCGGATCAACCGAGTCTC 3382  
Db 1319 GATCTCTGACACCGAGAAACAGAGCAGAGGCGAGACCGATCAGCGGATCAACCGAGTCTC 1378  
QY 3383 GCGAGTGTGAGAGTGTGAGAGTGAAGAGACAAAGGCTGAGATTTCTCGCTCTGACAGCAG 3442  
Db 1379 GCGAGTGTGAGAGTGTGAGAGTGAAGAGACAAAGGCTGAGATTTCTCGCTCTGACAGCAG 1438  
QY 3443 CTCTCAAGAGCAGAAAGCTGAAGGCGAGAGCTCTGACCAAGTCAATGACCTGAGAG 3502  
Db 1439 CTCTCAAGAGCAGAAAGCTGAAGGCGAGAGCTCTGACCAAGTCAATGACCTGAGAG 1498  
QY 3503 AGAAGCATGCTATGCTTGAATGAATGCCGAAAGCTTACAGCAGAAAGCTGAGAGCTGAAAC 3562  
Db 1499 AGAAGCATGCTATGCTTGAATGAATGCCGAAAGCTTACAGCAGAAAGCTGAGAGCTGAAAC 1558  
QY 3563 GAGAGCTCAACAGAGGCTCTGAGAGAGAGCAAGCCAAATTACAGAGAGATGAGACCTGC 3622  
Db 1559 GAGAGCTCAACAGAGGCTCTGAGAGAGAGCAAGCCAAATTACAGAGAGATGAGACCTGC 1618  
QY 3623 AGAAAAATCAATTTTCCGTCTGATCTCAAGAGCTGCAAGAGCTCTAGATCGGCTGATC 3682  
Db 1619 AGAAAAATCAATTTTCCGTCTGATCTCAAGAGCTGCAAGAGCTCTAGATCGGCTGATC 1678  
QY 3683 TACTGAAGACAGAAAGAGTGAAGTCTGAGATGATGATGAGAAACATTCAGGTTCTTAAT 3742  
Db 1679 TACTGAAGACAGAAAGAGTGAAGTGAAGTGAATGATGATGAGAAACATTCAGGTTCTTAAT 1738  
QY 3743 CTGATGAAAAGGTGAAGAAATGGAAGGCACTATTTCTCAACAAACCAAACTCATTAATTTTC 3802  
Db 1739 CTGATGAAAAGGTGAAGAAATGGAAGGCACTATTTCTCAACAAACCAAACTCATTAATTTTC 1798  
QY 3803 TGCAGCCAAATATGACCAACTGCTTAAAAAGAAAAGTTCCTCTGACAGTCAATGAGC 3862  
Db 1799 TGCAGCCAAATATGACCAACTGCTTAAAAAGAAAAGTTCCTCTGACAGTCAATGAGC 1858  
QY 3863 TGAAGTGGCCCTGAGAGAGAGAAAGTCTGCTGTGAGAGCTAGAGAAAGCCCTTACA 3922  
Db 1859 TGAAGTGGCCCTGAGAGAGAGAAAGTCTGCTGTGAGAGCTAGAGAAAGCCCTTACA 1918  
QY 3923 AGACCCGCAATCGAGCTCCGCTCCGCGGAGAGAACTGCTCCACCCGAAAGCAAGCA 3982  
Db 1919 AGACCCGCAATCGAGCTCCGCTCCGCGGAGAGAACTGCTCCACCCGAAAGCAAGCA 1978  
QY 3983 ACCCAGCCATTCAGAGCCAGCCAGCCAGAGAGAGAGATGCGCATGTCCGATCGTGC 4042  
Db 1979 ACCCAGCCATTCAGAGCCAGCCAGCCAGAGAGAGATGCGCATGTCCGATCGTGC 2038  
QY 4043 GGTCCGCAAGAGACACAGCCAGTGCATGAGTCTGTGGCCCGCCATTCAGGCTGAGAA 4102  
Db 2039 GGTCCGCAAGAGACACAGCCAGTGCATGAGTCTGTGGCCCGCCATTCAGGCTGAGAA 2098  
QY 4103 AGAGAGCTTCAACTCCAGAGAAATTTAGTGGGCTTTAAGAGAACCATGCAACCAATA 4162  
Db 2099 AGAGAGCTTCAACTCCAGAGAAATTTAGTGGGCTTTAAGAGAACCATGCAACCAATA 2158  
QY 4163 TTCTTCACCGATTCACGTAAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 4222  
Db 2159 TTCTTCACCGATTCACGTAAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 2218  
QY 4223 ATACCGTGACCTTGAAGAGCCAGGATCCAAATGTCTGAAATGTCAAGTGTGTGTAC 4282  
Db 2219 ATACCGTGACCTTGAAGAGCCAGGATCCAAATGTCTGAAATGTCAAGTGTGTGTAC 2278  
QY 4283 CCAAGTGTCCAGCTGCTTGCAGAGCACTGAGGCTTGTGCTGAAATGTGACACACT 4342  
Db 2279 CCAAGTGTCCAGCTGCTTGCAGAGCACTGAGGCTTGTGCTGAAATGTGACACACT 2338  
QY 4343 TCACGAGGCTTGTGCTGCTGCAAAATGAATGAATGCCAGGTCTCCAGACCAAGAGCCCA 4402  
Db 2339 TCACGAGGCTTGTGCTGCTGCAAAATGAATGAATGCCAGGTCTCCAGACCAAGAGCCCA 2398  
QY 4403 GCAGAGCTTGCACCTGGAAGAGGTGATGAAGGTGCCAGGAATTAACAAACAGAGCAGC 4462  
Db 2399 GCAGAGCTTGCACCTGGAAGAGGTGATGAAGGTGCCAGGAATTAACAAACAGAGCAGC 2458  
QY 4463 AAGCTGGAAGAGAGTGAATGCTCTGAGAGGAGTCAAAAGTCCCTATTAAGCAATG 4522  
Db 2459 AAGCTGGAAGAGAGTGAATGCTCTGAGAGGAGTCAAAAGTCCCTATTAAGCAATG 2518  
QY 4523 AAGCCAGAAAGCTGAGACAGAGCCGCTGAGAAATTTGAGCTGTGCTTCCGAGCGGG 4582  
Db 2519 AAGCCAGAAAGCTGAGACAGAGCCGCTGAGAAATTTGAGCTGTGCTTCCGAGCGGG 2578  
QY 4583 ATGTATCATTCATGAGGCGCTTGTGCTTCCGAAATTCAGCAATTCAGCAATGAGATG 4642  
Db 2579 ATGTATCATTCATGAGGCGCTTGTGCTTCCGAAATTCAGCAATTCAGCAATGAGATG 2638  
QY 4643 TCCCATCATTCATGAGATGAGATCTACCCGACACCACTGCTGTGGCCGGAGAAACC 4702  
Db 2639 TCCCATCATTCATGAGATGAGATCTACCCGACACCACTGCTGTGGCCGGAGAAACC 2698  
QY 4703 TCTACTGTGATGCTCCAGGCTTCCCTGACAAACAGGCTGGGTCAAGCTTGAATCAG 4762



```

Db      |||||
2699  TCCTACTGTAGTCCAGCTCCCTGACAAACAGCGCTGGGTACCGCTTGATGATCAG 2758
Qy      |||||
4763  TTGTCCGAGGTGGGAGAGTTTCTAGGAAAAAGAGAAAGTGAATGCTTAACTCTTGGA 4822
Db      |||||
2759  TTGTCCGAGGTGGGAGAGTTTCTAGGAAAAAGAGAAAGTGAATGCTTAACTCTTGGA 2818
Qy      |||||
4823  ACTCCCTGTGAAACTGGAAGGTGATACCGTCTAGACATGAATGACGCTCCCTTCA 4882
Db      |||||
2819  ACTCCCTGTGAAACTGGAAGGTGATACCGTCTAGACATGAATGACGCTCCCTTCA 2878
Qy      |||||
4883  GTGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4942
Db      |||||
2879  GTGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2938
Qy      |||||
4943  ACTCCCTTAACCATGCTCCGAGAAATGAGACATGCTTCAATTTATTTATTAAGAGAC 5002
Db      |||||
2939  ACTCCCTTAACCATGCTCCGAGAAATGAGACATGCTTCAATTTATTTATTAAGAGAC 2998
Qy      |||||
5003  TGGAGAAAGCTAATCATATATAGCAGAGAAAGCGGGCACTGTGTCTTGTGTGAAGTA 5062
Db      |||||
2999  TGGAGAAAGCTAATCATATATAGCAGAGAAAGCGGGCACTGTGTCTTGTGTGAAGTA 3058
Qy      |||||
5063  AAGTGAACAGTCCCTGGGCGGACCTGCTGCGGAGCGGCGGACATTCACCGACA 5122
Db      |||||
3059  AAGTGAACAGTCCCTGGGCGGACCTGCTGCGGAGCGGCGGACATTCACCGACA 3118
Qy      |||||
5123  TTTTGAAGCTGTCAAGGGCTGACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5182
Db      |||||
3119  TTTTGAAGCTGTCAAGGGCTGACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3178
Qy      |||||
5183  GCATCTGTGAGCCATGCCAGCAAAAGTGTCTATTCTCCGCTACAAAGAAACCTTACA 5242
Db      |||||
3179  GCATCTGTGAGCCATGCCAGCAAAAGTGTCTATTCTCCGCTACAAAGAAACCTTACA 3238
Qy      |||||
5243  AATCTGATCCGGAAAGAGATGAGACTCAGAGCCCTGACATGATCACTTGACCA 5302
Db      |||||
3239  AATCTGATCCGGAAAGAGATGAGACTCAGAGCCCTGACATGATCACTTGACCA 3298
Qy      |||||
5303  ATTACAGTATCTCTATTGGAACCAATTAATTTTACGAATGACATGAAGCAGTACACG 5362
Db      |||||
3299  ATTACAGTATCTCTATTGGAACCAATTAATTTTACGAATGACATGAAGCAGTACACG 3358
Qy      |||||
5363  TCGAGGAATTCCTGGATTAAGATGACATTCCTGAGACCTGTGTGTGTGTGTGTGTGT 5422
Db      |||||
3359  TCGAGGAATTCCTGGATTAAGATGACATTCCTGAGACCTGTGTGTGTGTGTGTGTGT 3418
Qy      |||||
5423  CCAACAGCTTCCTGTCTCAATCGTGAAGTGAACGCGAGCGGACGAGAGAGTACT 5482
Db      |||||
3419  CCAACAGCTTCCTGTCTCAATCGTGAAGTGAACGCGAGCGGACGAGAGAGTACT 3478
Qy      |||||
5483  TGCCTGTCTCCACGAAATTTGGAGTGTCTGTGAATTTTACGGAAGACGTAGCGACAG 5542
Db      |||||
3479  TGCCTGTCTCCACGAAATTTGGAGTGTCTGTGAATTTTACGGAAGACGTAGCGACAG 3538
Qy      |||||
5543  ACGATCTCAAGTGAAGTGGCTTACCTTTGCTTGGCTTGCCTACAGAAACCTTATCTTTG 5602
Db      |||||
3539  ACGATCTCAAGTGAAGTGGCTTACCTTTGCTTGGCTTGCCTACAGAAACCTTATCTTTG 3598
Qy      |||||
5603  TGAACCACTTCAACTCACTGGAAGTAATTGAATCCAGGCAAGCTCTCTACAGAGGACCC 5662
Db      |||||
3599  TGAACCACTTCAACTCACTGGAAGTAATTGAATCCAGGCAAGCTCTCTACAGAGGACCC 3658
Qy      |||||
5663  CTGCGGAGGCTGACTGGAATCCCGAACCCGCGCTACCTGGGCGCTGCAATTTCTCTAG 5722
Db      |||||
3659  CTGCGGAGGCTGACTGGAATCCCGAACCCGCGCTACCTGGGCGCTGCAATTTCTCTAG 3718
Qy      |||||
5723  GAGGATTTAATTGGGCTCTCTATACCAAGATTAATTAAGGCTCATTTGTGTGAAGAA 5782
Db      |||||
3719  GAGGATTTAATTGGGCTCTCTATACCAAGATTAATTAAGGCTCATTTGTGTGAAGAA 3778
Qy      |||||
5783  ACCCTGTGAAGAGTCCGAGCTGAACACACCGGGGCGCTGTCACTCCCGAGAGAGCC 5842

```

```

Db      |||||
3779  ACCTGTGAAGAGTCCGGCACTGAACACACCGGGGCGCTGCCACTCCGACAGAGCC 3838
Qy      |||||
5843  CCAACACAGAGGCGCCACCTCCAGTACACAGACACATCAACCAAGCGCTGCTTCCAGC 5902
Db      |||||
3839  CCAACACAGAGGCGCCACCTCCAGTACACAGACACATCAACCAAGCGCTGCTTCCAGC 3898
Qy      |||||
5903  CAGCGCGCGCGAAGGCGCCAGACACCGCGAGAGCCAAAGACACCCACCGCTACCGCG 5962
Db      |||||
3899  CAGCGCGCGCGAAGGCGCCAGACACCGCGAGAGCCAAAGACACCCACCGCTACCGCG 3958
Qy      |||||
5963  AGGGCGGAGACCGAGTGTGCGACAGGACAAAGTCTCTGCGCGCGCTGAGAGAGAAAGT 6022
Db      |||||
3959  AGGGCGGAGACCGAGTGTGCGACAGGACAAAGTCTCTGCGCGCGCTGAGAGAGAAAGT 4018
Qy      |||||
6023  CCGCGCGCGGATATCTACGACCGGAGAGAGCGGTCCCGCGACAGGCTGTTTGAAGCA 6082
Db      |||||
4019  CCGCGCGCGGATATCTACGACCGGAGAGAGCGGTCCCGCGAGGCTGTTTGAAGCA 4078
Qy      |||||
6083  GCAGCAGGGGCGGCTGCTGCGGAGCGGAGCCGCTGTCCAGGTGAACAAAG 6142
Db      |||||
4079  GCAGCAGGGGCGGCTGCTGCGGAGCGGAGCCGCTGTCCAGGTGAACAAAG 4138
Qy      |||||
6143  TCTGGACCACTTCTTCAATTA 6165
Db      |||||
4139  TCTGGACCACTTCTTCAATTA 4161

RESULT 12
ABR08361
ID  ABR08361 standard; cDNA, 3131 BP.
AC  ABA08361;
DT  11-JAN-2002 (first entry)
DE  Human RHO/RAC effector homologue-encoding cDNA. SEQ ID NO:137.
XX
XX
XX  Human; cytokine; cell proliferation; cell differentiation; growth factor;
XX  haematopoiesis regulation; tissue growth; immunomodulator; activin;
XX  inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
XX  proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX  myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
XX  chronic inflammatory condition; proliferative retinopathy;
XX  atherosclerosis; coronary heart disease; arterial ischaemia;
XX  bone disorder; osteoporosis; vascular growth disorder;
XX  tissue regeneration; wound healing; infection; immune disorder;
XX  cell culture; drug screening; gene therapy; antiinflammatory;
XX  antiaslarmic; antiarthritic; haemostatic; antiarteriosclerotic;
XX  cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
XX  antifungal; vulnery; antitumor; ss.
XX
XX  Homo sapiens.
XX
XX  WO200157188-A2.
XX
XX  09-AUG-2001.
XX
XX  05-FEB-2001; 2001WO-US03800.
XX
XX  03-FEB-2000; 2000US-0496914.
XX  27-APR-2000; 2000US-0560875.
XX
XX  (HYS- ) HYSBQ INC.
XX
XX  Tang YT, Liu C, Dymnac RT;
XX
XX  WPI; 2001-457740/49.
XX  P-PDB; ABB11117.
XX
XX  Human proteins and DNA encoding sequences useful for preventing,
XX  treating or ameliorating a medical condition in a mammalian subject
XX  e.g. arthritis and cancer -
XX

```





QY	4405	AGCAGCTTCACCTGGAAAGGTGATGAAGGTGCCAGGAATTAACAACGAGGACAGCA	4464
Db	1501	AACAGCTTCACCTGGAAAGGTGATGAAGGTGCCAGGAATTAACAACGAGGACAGCA	1566
QY	4465	GGCTGGGACAGGAAGTACATTGTCCGGAGGGATCAAAAAGTCTCATTTATTAACAATTGA	4524
Db	1561	GGCTGGGACAGGAAGTACATTGTCCGGAGGGATCAAAAAGTCTCATTTATTAACAATTGA	1620
QY	4525	GCCAGGAAGCTGGA CAGAGGCCGGTGGAAAGATTGAGCTGTGCTTCCGA CGGGAT	4584
Db	1621	GCCAGGAAGCTGGA CAGAGGCCGGTGGAAAGATTGAGCTGTGCTTCCGA CGGGAT	1680
QY	4585	GTATCTATTCA TGTGTCGGTGTGTCTTCCGA CTGCGAAATACAGCCAAAGCAGATGTC	4644
Db	1681	GTATCTATTCA TGTGTCGGTGTGTCTTCCGA CTGCGAAATACAGCCAAAGCAGATGTC	1740
QY	4645	CCATTCATACATCGAAGATGGAATTCACCCGCA CACCACTGCTGGCCCGGAGAACCTTC	4704
Db	1741	CCATTCATACATCGAAGATGGAATTCACCCGCA CACCACTGCTGGCCCGGAGAACCTTC	1800
QY	4705	TACTTGTAGCTCCGAGCTCCCTTGACAAACAGCGCTGGGTCA CGGCTTTGAATCAAGTT	4764
Db	1801	TACTTGTAGCTCCGAGCTCCCTTGACAAACAGCGCTGGGTCA CGGCTTTGAATCAAGTT	1860
QY	4765	GTGCGAGGTGGGAGATTTCTPAGGAAACGAGAGCTAGTGCTAAACCTGCTTGGAAAC	4824
Db	1861	GTGCGAGGTGGGAGATTTCTPAGGAAACGAGAGCTAGTGCTAAACCTGCTTGGAAAC	1920
QY	4825	TCCCGCTGAAACCTGGAAGGTGATGACCGGTAGACATGAACTGCACGCTGCCCTTCAAGT	4884
Db	1921	TCCCGCTGAAACCTGGAAGGTGATGACCGGTAGACATGAACTGCACGCTGCCCTTCAAGT	1980
QY	4885	GACCAGGTGTGTGTGGTGGGACACGAGGAAGGCTCTPACGCGCTGAATGTCTTGAAGAAC	4944
Db	1981	GACCAGGTGTGTGTGGTGGGACACGAGGAAGGCTCTPACGCGCTGAATGTCTTGAAGAAC	2040
QY	4945	TCCCTAACCCATGTCCACGAAATTGGAGCACTTTCNAATTTATATTATCAAGSACTG	5004
Db	2041	TCCCTAACCCATGTCCACGAAATTGGAGCACTTTCNAATTTATATTATCAAGSACTG	2100
QY	5005	GAGAAAGTACATCATGATATGACGAGGAAGAGGGGCACTGTGTCTTGTGGACGTGAAGAA	5064
Db	2101	GAGAAAGTACATCATGATATGACGAGGAAGAGGGGCACTGTGTCTTGTGGACGTGAAGAA	2160
QY	5065	GTGAAACAGTCCCTGCGCCCACTGCCACCTGCTGCCAGGCCCGACATCTCACCCAACTT	5124
Db	2161	GTGAAACAGTCCCTGCGCCCACTGCCACCTGCTGCCAGGCCCGACATCTCACCCAACTT	2220
QY	5125	TTTGAAGCTGCAAGGGCTGCGCACTGTGTTGGGGCAGGCAAGATTGGAACCGGGCTCTGC	5184
Db	2221	TTTGAAGCTGCAAGGGCTGCGCACTGTGTTGGGGCAGGCAAGATTGGAACCGGGCTCTGC	2280
QY	5185	ATCTGTGAGGCATAGCCACGAAAGTGTGCTATTCTCGCTACACGAAACCTTAGCAAA	5244
Db	2281	ATCTGTGAGGCATAGCCACGAAAGTGTGCTATTCTCGCTACACGAAACCTTAGCAAA	2340
QY	5245	TACTGCAATCCGGAAGAGATGAGACCTCAGAGCCCTGCAAGCTGATCCACTTCAACAAAT	5304
Db	2341	TACTGCAATCCGGAAGAGATGAGACCTCAGAGCCCTGCAAGCTGATCCACTTCAACAAAT	2400
QY	5305	TACAGTATCTCATTTGGAACCAATAAATTCTACGAAATGACATGAAGCAGTACAGCTTC	5364
Db	2401	TACAGTATCTCATTTGGAACCAATAAATTCTACGAAATGACATGAAGCAGTACAGCTTC	2460
QY	5365	GAGGAATTCCTCGATTAAGATGACATTTCTTGGCACTGTGCTGTGTTGCGGCTCTTTC	5424
Db	2461	GAGGAATTCCTCGATTAAGATGACATTTCTTGGCACTGTGCTGTGTTGCGGCTCTTTC	2520
QY	5425	AACAGCTTCCTGTCTCAATCGTGCAGGTGAACAGCGCAGGGACGAGGAGGAGTACTTG	5484
Db	2521	AACAGCTTCCTGTCTCAATCGTGCAGGTGAACAGCGCAGGGACGAGGAGGAGTACTTG	2580
QY	5485	CTGTGTTTCCACGAATTTGGAGTGTTCGTGATTTCTTACGGAAGAGTACGCCGACAGAC	5544

Db 2581 CTGTTTCCACGAATTTGAGTGTCGTGGATTCTTACGGAAACGTAAGCCGCACAGAC 2640

QY 5545 GATCTCAGTGAAGTCGCTTACCTTTGGCCCTTGGCTACAGAGAACCTTATCTGTTTGTG 5604

Db 2641 GATCTCAAGTGAAGTCGCTTACCTTTGGCCCTTGGCTACAGAGAACCTTATCTGTTTGTG 2700

QY 5605 ACCCACTTCAACTACTGGAAGTATTGAGATCCAGGACACGCTCTCAGCAGGAGACCCCT 5664

Db 2701 ACCCACTTCAACTACTGGAAGTATTGAGATCCAGGACACGCTCTCAGCAGGAGACCCCT 2760

QY 5665 GCCCGAGGTAACCTGAGATCTCCGAAACCCGCGCTAAGCTGGGCGCTGCCATTTCTCTCAGGA 5724

Db 2761 GCCCGAGGTAACCTGAGATCTCCGAAACCCGCGCTAAGCTGGGCGCTGCCATTTCTCTCAGGA 2820

QY 5725 GCGAATTTACTGGCGTCCCTCATACACAGATTTAAATTAAGGGTCATTTCGTGCAAGGAAAC 5784

Db 2821 GCGAATTTACTGGCGTCCCTCATACACAGATTTAAATTAAGGGTCATTTCGTGCAAGGAAAC 2880

QY 5785 CTCTGGAAGAGTCCCGCATCTGAACACACACCGGGGGCCGCTCCACTCTCCGAG 5837

Db 2881 CTCTGGAAGAGTCCCGCATCTGAACACACACCGGGGGCCGCTCCACTCTCCGAG 2933

CC	XX	result 13
CC	XX	AAL55216
CC	XX	ID AAL55216 standard; DNA; 5261 BP.
CC	XX	AC AAL55216;
CC	XX	DT 01-MAY-2003 (first entry)
CC	XX	DE Human CR1K related DNA sequence, SEQ ID No 6.
CC	XX	KW Anorectic; hypotensive; cardiant; antilipaeamic; cerebroprotective;
CC	XX	KW antiagut; osteopathic; antiarthritic; cytoslatic; antidepressant;
CC	XX	KW immunomodulator; antihemic; tranquilliser; antiparkinsonian; nootropic;
CC	XX	KW neuroprotective; antinflammatory; antidiabetic; analgesic;
CC	XX	KW human citron rho/rac-interacting kinase; enzyme; CR1K; ameliorating;
CC	XX	KW obesity; comorbidity; cancer; anorexia; cachexia; bulimia;
CC	XX	KW central nervous system disorder; chronic obstructive pulmonary disease;
CC	XX	KW diabetes; pain; ds.
CC	XX	OS Homo sapiens.
CC	XX	PV WOZ003004523-A1.
CC	XX	ED 16-JAN-2003.
CC	XX	PF 28-JUN-2002; 2002WO-EP07156.
CC	XX	PR 02-JUL-2001; 2001US-301841P.
CC	XX	PR 11-DEC-2001; 2001US-338651P.
CC	XX	PR 25-APR-2002; 2002US-375014P.
CC	XX	PA (FARB ) BAYER AG.
CC	XX	PI Zhu Z;
CC	XX	WP1; 2003-221576/21.
CC	XX	PT New human citron rho/rac-interacting kinase (CR1K) polypeptide and
CC	XX	PT polynucleotide, useful in preventing, ameliorating or treating diseases
CC	XX	PT associated with human CR1K dysfunction, e.g. obesity, diabetes or
CC	XX	PT Alzheimer's disease -
CC	XX	PS Disclosure; Fig 6; 237pp; English.
CC	XX	The invention relates to an isolated polynucleotide encoding a human
CC	XX	citron rho/rac-interacting kinase polypeptide. The isolated
CC	XX	polynucleotide comprises a 6165 or 8603 base pair sequence, given in the
CC	XX	specification. The human citron rho/rac-interacting kinase (CR1K)
CC	XX	polypeptide and polynucleotide are useful in preventing, ameliorating, or

CC treating diseases associated with human CRK dysfunction such as obesity  
CC and obesity-associated comorbidities (e.g. hypertension, coronary artery  
CC disease, hyperlipidaemia, stroke, gout, osteoarthritis, some types of  
CC cancer including endometrial, breast, prostate and colon cancer),  
CC anorexia, cachexia, bulimia, central nervous system disorders (e.g. mood  
CC disorders, anxiety disorders, Parkinson's disease or Alzheimer's  
CC disease), chronic obstructive pulmonary disease, or diabetes. These can  
CC also be used to treat pain associated with the disorders. The human CRK  
CC polypeptide is also useful in diagnostic assays or in genetic testing.  
CC The expression vector or the reagent is useful in preparing a medicament  
CC for modulating the activity of a human CRK in a disease, e.g. obesity, a  
CC central nervous system disorder, or chronic obstructive pulmonary  
CC disease. The fusion protein is useful for generating antibodies against a  
CC CRK polypeptide and for use in various assay systems. The methods are  
CC useful in producing and detecting the polynucleotide and polypeptide and  
CC in screening for agents that modulate the activity of the human CRK  
CC polypeptide. This polynucleotide sequence represents a DNA sequence  
CC relating to the human CRK protein of the invention.  
CC  
XX

Sequence 5261 BP; 1309 A; 1465 C; 1283 G; 1204 T; 0 other;

Query Match 45.7%; Score 2818.2; DB 25; Length 5261;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2820; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 3343 CAAAGCAGGCGGAGAGCCGATCAGCGGATCAGAGCTTCCGCAAGTGTGTGAGCTGGCA 3402
Db 1 CAAGAGCAGGCGGAGAGCCGATCAGCGGATCAGAGCTTCCGCAAGTGTGTGAGCTGGCA 60
QY 3403 GTAGAGAGCAGCAGGCTGAGATTCTCGCTCAGAGGCTCTCAAGAGCGAAGAGCTG 3462
Db 61 GTAGAGAGCAGCAGGCTGAGATTCTCGCTCAGAGGCTCTCAAGAGCGAAGAGCTG 120
QY 3463 AAGCGCAGAGCTCTCTGCAAGACTCAATGACTGAGAGAGAGAGCTGATGCTTGA 3522
Db 121 AAGCGCAGAGCTCTCTGCAAGACTCAATGACTGAGAGAGAGAGCTGATGCTTGA 180
QY 3523 ATGAAATGCCGAAAGCTTACGCGAAGCTGGAAGCTGAGAGAGAGCTCAAGAGAGCTT 3582
Db 181 ATGAAATGCCGAAAGCTTACGCGAAGCTGGAAGCTGAGAGAGAGCTCAAGAGAGCTT 240
QY 3583 CTGGAAGAGCAGCAGAAATTACAGCAGAGATGAGACCTGAGAAATATCATTTCCGT 3642
Db 241 CTGGAAGAGCAGCAGAAATTACAGCAGAGATGAGACCTGAGAAATATCATTTCCGT 300
QY 3643 CTGACTCAAGGACTGCAAGAGCTCTAGATCGGGCTGATCTACTGGAAGCAGAAAGAGT 3702
Db 301 CTGACTCAAGGACTGCAAGAGCTCTAGATCGGGCTGATCTACTGGAAGCAGAAAGAGT 360
QY 3703 GACTTGGAGTATCAGCTGGAAGAAATTCAAGCTTCTATTTCTCATGAAAAGTGAAATG 3762
Db 361 GACTTGGAGTATCAGCTGGAAGAAATTCAAGCTTCTATTTCTCATGAAAAGTGAAATG 420
QY 3763 GAAGGACTATTTCTCAAGAAACCAACTCATGATTTTCTGCAAGCCAAATGAGCA 3822
Db 421 GAAGGACTATTTCTCAAGAAACCAACTCATGATTTTCTGCAAGCCAAATGAGCA 480
QY 3823 CTTGCTAAAAAGAAAGGTTCTCTGCAATCAATGAGCTGAAAGTGGCCCTGAGAAAG 3882
Db 481 CTTGCTAAAAAGAAAGGTTCTCTGCAATCAATGAGCTGAAAGTGGCCCTGAGAAAG 540
QY 3883 GAGAAGCTCGCTGTGCAAGGCTAGAGGAAGCCCTTCAAGAAACCCGATTCGAGCTCCGG 3942
Db 541 GAGAAGCTCGCTGTGCAAGGCTAGAGGAAGCCCTTCAAGAAACCCGATTCGAGCTCCGG 600
QY 3943 TCCGCCCGGAGAGAGCTCCCAACGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 4002
Db 601 TCCGCCCGGAGAGAGCTCCCAACGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 660
QY 4003 GGCACCGCAGAGCAGAGATGCGCATGTCGCCATGTGCGGTGCGCAGAGCAAGCAAGCC 4062
Db 661 GGCACCGCAGAGCAGAGATGCGCATGTCGCCATGTGCGGTGCGCAGAGCAAGCAAGCC 720
```

```
QY 4063 AGTCCATGAGCTGTGAGGCCCGCCATCCAGCCGAGAAAGAGTCTTAACCTCCAGAG 4122
Db 721 AGTCCATGAGCTGTGAGGCCCGCCATCCAGCCGAGAAAGAGTCTTAACCTCCAGAG 780
QY 4123 GAATTTAGTGGGCTCTTAAGAAACGATCAGCAACCAATATTTCTCCAGATTCAGCTA 4182
Db 781 GAATTTAGTGGGCTCTTAAGAAACGATCAGCAACCAATATTTCTCCAGATTCAGCTA 840
QY 4183 GAGCTGAACATGAGAGCAAAAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4242
Db 841 GAGCTGAACATGAGAGCAAAAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900
QY 4243 CAGGATCCAAATGTCTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 4302
Db 901 CAGGATCCAAATGTCTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 960
QY 4303 CCAGCCACTCGGGCTTGGCTGCTGAATATGCAACACTTCAACCGAGGCTTTCGCTG 4362
Db 961 CCAGCCACTCGGGCTTGGCTGCTGAATATGCAACACTTCAACCGAGGCTTTCGCTG 1020
QY 4363 GACAAATGAATCCCCAGGCTCCAGACCAAGAGCCGAGCAGACCTTGACCTGAGAA 4422
Db 1021 GACAAATGAATCCCCAGGCTCCAGACCAAGAGCCGAGCAGACCTTGACCTGAGAA 1080
QY 4423 GGGTGAATGAAGGTGCCAGAGATTAACAAAGAGCAGCAAGGCTGGAGCAGAGATAC 4482
Db 1081 GGGTGAATGAAGGTGCCAGAGATTAACAAAGAGCAGCAAGGCTGGAGCAGAGATAC 1140
QY 4483 ATTGCTCTGAGAGGATCAAAAGTCTCTATTTATGACATGTAAGCCAGAGAGCTGACAG 4542
Db 1141 ATTGCTCTGAGAGGATCAAAAGTCTCTATTTATGACATGTAAGCCAGAGAGCTGACAG 1200
QY 4543 AGCCGAGTGAAGAAATTTAGAGTGTGCTCCGAGGAGGAGTATATTCATGTGAGCC 4602
Db 1201 AGCCGAGTGAAGAAATTTAGAGTGTGCTCCGAGGAGGAGTATATTCATGTGAGCC 1260
QY 4603 GTTGTGCTTCCGAACTCGCAATTAACAGCAAGAGAGATGTCCCATACATACAGAGATG 4662
Db 1261 GTTGTGCTTCCGAACTCGCAATTAACAGCAAGAGAGATGTCCCATACATACAGAGATG 1320
QY 4663 GAATTCACCCGACACACCTGTGCTGGCCCGGAGAAACCTCTGATCTTGTAGCTCCAGC 4722
Db 1321 GAATTCACCCGACACACCTGTGCTGGCCCGGAGAAACCTCTGATCTTGTAGCTCCAGC 1380
QY 4723 TTCCTCGAAGAAACAGGCTGAGGCTCACCGCTTAAGATGTGTGCGAGGTGGAGAGTT 4782
Db 1381 TTCCTCGAAGAAACAGGCTGAGGCTCACCGCTTAAGATGTGTGCGAGGTGGAGAGTT 1440
QY 4783 TCTAGGAAAGAGCAGAGAGCTGATGCTAAATGCTTGAAGCTCCCTGTGAAGCTGAA 4842
Db 1441 TCTAGGAAAGAGCAGAGAGCTGATGCTAAATGCTTGAAGCTCCCTGTGAAGCTGAA 1500
QY 4843 GGTGATGACCGCTTGAACATGAATGCAAGCTGCTGCTTCAAGTGAAGAGGTGTGTGTG 4902
Db 1501 GGTGATGACCGCTTGAACATGAATGCAAGCTGCTGCTTCAAGTGAAGAGGTGTGTGTG 1560
QY 4903 GGCACCGAAGAGGCTTACGCGCTGAATGTCTTGAAGAACTCCCTAAACCAATGCCA 4962
Db 1561 GGCACCGAAGAGGCTTACGCGCTGAATGTCTTGAAGAACTCCCTAAACCAATGCCA 1620
QY 4963 GGAATTTGAGCAGTCTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
Db 1621 GGAATTTGAGCAGTCTTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
QY 5023 GGAAGAGAGAGAGGAGCTGTGTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5082
Db 1681 GGAAGAGAGAGAGGAGCTGTGTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1740
QY 5083 CAGTCCCACTGCTGCGCAGCCGACATCTCAACCAACATTTTGAAGCTGTCAAGAGG 5142
Db 1741 CAGTCCCACTGCTGCGCAGCCGACATCTCAACCAACATTTTGAAGCTGTCAAGAGG 1800
QY 5143 TGCCACTTGTGTTGGGCGAGGCAAGATTGAAGAGGAGCTGTGATCTGTGACAGCATGCC 5202
```

```

Db      1801  TGCCACTTGTGGGGGAGGAGATTGAGAACGGGCTGTGCATCTGTGAGCCATCCC 1860
QY      5203  AGCAAAAGTCATTCCTCCGCTACAGGAAAACCTGAGCAATCTGCATCTCCGAAAAG 5262
Db      1861  AGCAAAAGTCATTCCTCCGCTACAGGAAAACCTGAGCAATCTGCATCTCCGAAAAG 1920
QY      5263  ATAGAGCCTCAGAGCCCTGAGCTGTATCACTTCAACCAATTCATATCTCATTTGA 5322
Db      1921  ATAGAGCCTCAGAGCCCTGAGCTGTATCACTTCAACCAATTCATATCTCATTTGA 1980
QY      5323  ACCAATAATTTCTAGAAATGCAATGAGCATGACGCTCGAGAAATTCCTGATTAAG 5382
Db      1981  ACCAATAATTTCTAGAAATGCAATGAGCATGACGCTCGAGAAATTCCTGATTAAG 2040
QY      5383  AATGACCATTCCTTGGGACCTGCTGTGTTTGGCCCTTCCAGACCTCCGCTGCA 5442
Db      2041  AATGACCATTCCTTGGGACCTGCTGTGTTTGGCCCTTCCAGACCTCCGCTGCA 2100
QY      5443  ATCGTGAAGTGAACAGCGACAGGAGGAGAGAGTACTGTGTTCACGAATTT 5502
Db      2101  ATCGTGAAGTGAACAGCGACAGGAGGAGAGTACTGTGTTCACGAATTT 2160
QY      5503  GGAAGTTCCTGGAATTTCTTACGGAAGAGTACGCGGACAGAGATCAAGTGAAGTGC 5562
Db      2161  GGAAGTTCCTGGAATTTCTTACGGAAGAGTACGCGGACAGAGATCAAGTGAAGTGC 2220
QY      5563  TTACCTTGGCCTTGTGCTACAGAAACCTATCTGTTGTGACCACTTCAACTGCTC 5622
Db      2221  TTACCTTGGCCTTGTGCTACAGAAACCTATCTGTTGTGACCACTTCAACTGCTC 2280
QY      5623  GAAATTAATTGATTCAGAGGACGCTCTCAGACAGGACCTTGCCTCGAGCTTCTGAC 5682
Db      2281  GAAATTAATTGATTCAGAGGACGCTCTCAGACAGGACCTTGCCTCGAGCTTCTGAC 2340
QY      5683  ATCCGGAACCGCGCTACCTGAGGCTGCTGCTTCTCAGAGAGGATTTCTTGGCTGC 5742
Db      2341  ATCCGGAACCGCGCTACCTGAGGCTGCTGCTTCTCAGAGAGGATTTCTTGGCTGC 2400
QY      5743  TCATACAGAGATTAATTAAGGCTATTTGCTGCAAGGAAAACCTGTGAAGAGTCCGC 5802
Db      2401  TCATACAGAGATTAATTAAGGCTATTTGCTGCAAGGAAAACCTGTGAAGAGTCCGC 2460
QY      5803  ACTGAACACCAACCGGCGCGCTGCACTCCGCGAGAGGCGCCCAACAGGAGGCGCAC 5862
Db      2461  ACTGAACACCAACCGGCGCGCTGCACTCCGCGAGAGGCGCCCAACAGGAGGCGCAC 2520
QY      5863  ACGTACAAAGCAATACCAAGCGCTGCTGCAAGCCGAGCGCGCGCGAGAGGCGCC 5922
Db      2521  ACGTACAAAGCAATACCAAGCGCTGCTGCAAGCCGAGCGCGCGCGAGAGGCGCC 2580
QY      5923  AGCCACCCCGGAGAGCCAAAGACACCCCAACCGCTACCGCGAGGCGCGAGCTGCGC 5982
Db      2581  AGCCACCCCGGAGAGCCAAAGACACCCCAACCGCTACCGCGAGGCGCGAGCTGCGC 2640
QY      5983  AGGGAACAAGTCTCTGCGCGCGCGCTGAGGAGAGAGTCCCGCGCGGATCTAGC 6042
Db      2641  AGGGAACAAGTCTCTGCGCGCGCGCTGAGGAGAGAGTCCCGCGCGGATCTAGC 2700
QY      6043  ACGCGAGAGAGCGGCTCCCGCGAGGCTGTTTGAAGACAGACAGAGGCGCGGCTGCT 6102
Db      2701  ACGCGAGAGAGCGGCTCCCGCGAGGCTGTTTGAAGACAGACAGAGGCGCGGCTGCT 2760
QY      6103  GCGGAGAGCGGAGAGAGCGGCTGTTCCAGAGTGAACAAGCTGAGGACCACTTCACTA 6162
Db      2761  GCGGAGAGCGGAGAGAGCGGCTGTTCCAGAGTGAACAAGCTGAGGACCACTTCACTA 2820
QY      6163  TAA 6165
Db      2821  TAA 2823

```

RESULT 14

```

ABV30132/c
ID      ABV30132 standard; cDNA; 2896 BP.
XX
AC      ABV30132;
XX
DT      16-SEP-2002 (first entry)
XX
DE      Human prostate expression marker cDNA 30123.
XX
KW      Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX      pharmacogenomic marker; gene; ss.
XX
OS      Homo sapiens.
XX
PN      M0200160860-A2.
XX
PD      23-AUG-2001.
XX
PF      20-FEB-2001; 2001WO-US05171.
XX
PR      17-FEB-2000; 2000US-183319P.
XX
PR      16-MAR-2000; 2000US-189862P.
XX
PR      25-MAY-2000; 2000US-207454P.
XX
PR      09-JUN-2000; 2000US-211314P.
XX
PR      18-JUL-2000; 2000US-219007P.
XX
PR      13-DEC-2000; 2000US-255281P.
XX
PA      (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI      Schlegel R, Endege WO, Monahan JB;
XX
DR      WPI, 2001-662795/76.
XX
PT      Novel isolated nucleic acid molecule associated with cancerous state of
PT      prostate cells and correlating with presence of prostate cancer, useful
PT      for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS      Claim 1, Page 6521-6522, 11750pp; English.
XX
XX
CC      The invention relates to an isolated nucleic acid molecule (1) comprising
CC      a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC      specification or its complement. (1) is useful for:
CC      (a) assessing whether a patient is afflicted with prostate cancer;
CC      (b) monitoring the progression of prostate cancer in a patient;
CC      (c) assessing the efficacy of a test compound to inhibit prostate
CC      cancer in a patient;
CC      (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC      in a patient;
CC      (e) selecting a composition for inhibiting prostate cancer in a patient;
CC      (f) assessing the prostate cell carcinogenic potential of a compound;
CC      (g) determining whether prostate cancer has metastasized in a patient;
CC      (h) assessing the aggressiveness or indolence of prostate cancer in a
CC      patient;
CC      (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ      Sequence 2896 BP; 525 A; 839 C; 647 G; 885 T; 0 other;
XX
Query Match      43.5%; Score 2682.8; DB 23; Length 2896;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 2825; Conservative 0; Mismatches 2; Indels 111; Gaps 3;
QY      744  AAACAAGATGATGATGACCAAACTCCGAGTGTGGAGCCCGAGATTCATGAGCTCTGAAGT 803
Db      2896  AAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2837
QY      804  GCTGACTGTGATGAACGCGGATGAGAAAGCACTTACGCGCTGAGCTGTGACTGTGTGTC 863
Db      2836  GCTGACTGTGATGAACGCGGATGAGAAAGCACTTACGCGCTGAGCTGTGACTGTGTGTC 2777
QY      864  AGTGGCGGTGATGCTTATGAGATGATTTATGAGGATCCCTTCGAGAGGAACTTC 923
Db      2776  AGTGGCGGTGATGCTTATGAGATGATTTATGAGGATCCCTTCGAGAGGAACTTC 2717

```

QY 924 TGCAGACCTTCAATTAACATTATGATTTTCACGGCTTTTGAATTTCCAGATGACC 983  
DB 2716 TGCAGAACCTTCAATTAACATTATGATTTTCACGGCTTTTGAATTTCCAGATGACC 2657  
QY 984 CAAAGTAGAGTGAATCTTTGATCTGATTCAAAGCTTGTGTGGCGCCAGAAAGAG 1043  
DB 2656 CAAAGTAGAGTGAATCTTTGATCTGATTCAAAGCTTGTGTGGCGCCAGAAAGAG 2597  
QY 1044 ACTGAATTTGAAGTCTTGTGCTGCCATCTTTCTTCTCTAAATTTGACTGAAACAT 1103  
DB 2596 ACTGAATTTGAAGTCTTGTGCTGCCATCTTTCTTCTCTAAATTTGACTGAAACAT 2537  
QY 1104 TCGTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1163  
DB 2536 TCGTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2477  
QY 1164 TGAATGACCAAGAGAAATTCGTGGGTTTCACTCTCTCTCTCTCTCTCTCTCTCTCT 1223  
DB 2476 TGAATGACCAAGAGAAATTCGTGGGTTTCACTCTCTCTCTCTCTCTCTCTCTCTCT 2417  
QY 1224 CTTCCTGGGTGAAGAACTGCGGTTTGTGGGTTTGTGTAACGCAAGGCACTGGGATCT 1283  
DB 2416 CTTCCTGGGTGAAGAACTGCGGTTTGTGGGTTTGTGTAACGCAAGGCACTGGGATCT 2357  
QY 1284 TGGTAAATCTGA----- 1295  
DB 2356 TGGTAAATCTGAATTTTGGGGGATGATGAATGATGCTTGAATGCTTGGTCTATGACC 2297  
QY 1296 -----GTCCTGTGTGTCTGGGCTCTGGACTCTCCCTGCAAGACTAGC 1335  
DB 2296 TGGCACAATGAAAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2237  
QY 1336 TCCATGGAAGAACTTCTCATCAAAAGCAAGCTTCAAGACTCTCAAGCAAGTGT 1395  
DB 2236 TCCATGGAAGAACTTCTCATCAAAAGCAAGCTTCAAGACTCTCAAGCAAGTGT 2177  
QY 1396 CACAAGATGAGAGAGAAATGACCGGTTTACATCGGAGAGTGTCAAGAGTGTGAGGCTGTG 1455  
DB 2176 CACAAGATGAGAGAGAAATGACCGGTTTACATCGGAGAGTGTCAAGAGTGTGAGGCTGTG 2117  
QY 1456 CTTAAGTCAGAGAGAGTGTGAGCTGAAAGCTCTGAGACTCAGAGATCCCTCCGCGAGCAG 1515  
DB 2116 CTTAAGTCAGAGAGAGTGTGAGCTGAAAGCTCTGAGACTCAGAGATCCCTCCGCGAGCAG 2057  
QY 1516 GACCTTGTCTACTCATCAAGATGCAATGACTTAAAGCGAATTTGAGAGCAAGCAG 1575  
DB 2056 GACCTTGTCTACTCATCAAGATGCAATGACTTAAAGCGAATTTGAGAGCAAGCAG 1997  
QY 1576 ATGAGAGTGTCCAGAGAGATGCAAGAGACTGAGACTCTCCATGATATTCAGAGAGCAG 1635  
DB 1996 ATGAGAGTGTCCAGAGAGATGCAAGAGACTGAGACTCTCCATGATATTCAGAGAGCAG 1937  
QY 1636 AGCCGGAAGCTCCAGAAATCAAAAGAGAGAGTACAGAGCTCAAGTGTGAAGAAATGAGG 1695  
DB 1936 AGCCGGAAGCTCCAGAAATCAAAAGAGAGAGTACAGAGCTCAAGTGTGAAGAAATGAGG 1877  
QY 1696 TTGATGATGAATCAATTTGGAAGAGGATCTTGTCTCAGCAAGAGAGAGTATCTTAC 1755  
DB 1876 TTGATGATGAATCAATTTGGAAGAGGATCTTGTCTCAGCAAGAGAGAGTATCTTAC 1817  
QY 1756 GAATCTGAGCTGAGAGAGTCTGGGCTTGTGCTGGAAGATTCAGAGGGAAGCCAGCAGAA 1815  
DB 1816 GAATCTGAGCTGAGAGAGTCTGGGCTTGTGCTGGAAGATTCAGAGGGAAGCCAGCAGAA 1757  
QY 1816 TGTCAAGCATAACTGTTGAAGGCTAAGATCAAGAGGAACTGAAAGTGGAGAAATATGCG 1875  
DB 1756 TGTCAAGCATAACTGTTGAAGGCTAAGATCAAGAGGAACTGAAAGTGGAGAAATATGCG 1697  
QY 1876 AAACGTGGAAGATGATGAATGCTGAGAGAGCTCAAAATTCAGAGAGTCTCAAGGAACTG 1935  
DB 1696 AAACGTGGAAGATGATGAATGCTGAGAGAGCTCAAAATTCAGAGAGTCTCAAGGAACTG 1637  
QY 1936 GAGAAAGCTGTAAAAAGCCAGACGAGAGGCAACGAGCTGTCTCAGATATTCGCGCAGGCA 1995

DB 1636 GAGAAAGCTGTAAAAAGCCAGACGAGAGGCCAGGACTGTGAGAAATATCCGCCAGGCA 1577  
QY 1996 AAGAGCGAGCCGAGAGGAGCTGAGAAAGCTGCAAGAACGAGAGATTTCTTGAAGGC 2055  
DB 1576 AAGAGCGAGCCGAGAGGAGCTGAGAAAGCTGCAAGAACGAGAGATTTCTTGAAGGC 1517  
QY 2056 ATCAGAAAGAGCTGTGAGAGCTGAGGAAACGCCGCTATCTCTGAGAAACAAGTAAAG 2115  
DB 1516 ATCAGAAAGAGCTGTGAGAGCTGAGGAAACGCCGCTATCTCTGAGAAACAAGTAAAG 1457  
QY 2116 AGACTAGAGCAATGAGAGCTGAGAAACAGACTGAGAGATCATCCAGCAAAATCC 2175  
DB 1456 AGACTAGAGCAATGAGAGCTGAGAAACAGACTGAGAGATCATCCAGCAAAATCC 1397  
QY 2176 CAAAGATCCAGAGATGCTGATTAATTTCTGAGACTGAGAGAAACATCGGAGGCC 2235  
DB 1396 CAAAGATCCAGAGATGCTGATTAATTTCTGAGACTGAGAGAAACATCGGAGGCC 1337  
QY 2236 CAAAGTCCAGCCAGAGACTGAGAGTGCCTGAAACAGAAAGAGCAGCACTATGA-AGA 2294  
DB 1336 CAAAGTCCAGCCAGAGACTGAGAGTGCCTGAAACAGAAAGAGCAGCACTATGAAGGGA 1277  
QY 2295 AAAGATTAAGTGTGAGCAATCAGATAAAGAAAGCTGTGCTGAGAGAGACACTGGA 2354  
DB 1276 AAAGATTAAGTGTGAGCAATCAGATAAAGAAAGCTGTGCTGAGAGAGACACTGGA 1217  
QY 2355 GAAATGATGAGAGAGCAGAGAGAGAGGCCATGAGAAAGGCAAAATTTCTGCGAACA 2414  
DB 1216 GAAATGATGAGAGAGCAGAGAGAGAGGCCATGAGAAAGGCAAAATTTCTGCGAACA 1157  
QY 2415 GAAGGCGATGATCAATGCTATGATTCAGAAATCAGATCCCTGGAAACAGAGTTGGGA 2474  
DB 1156 GAAGGCGATGATCAATGCTATGATTCAGAAATCAGATCCCTGGAAACAGAGTTGGGA 1097  
QY 2475 ACTGTCTGAAGCCATTAATCTTGCAAGCAATGACAGTCTTTTACCAGAAAGATGAA 2534  
DB 1096 ACTGTCTGAAGCCATTAATCTTGCAAGCAATGACAGTCTTTTACCAGAAAGATGAA 1037  
QY 2535 GGGCCAGAAAGAGATGATTTCTGAACTCAGGCAACGAAATTTTACCTGGAACACAGGC 2594  
DB 1036 GGGCCAGAAAGAGATGATTTCTGAACTCAGGCAACGAAATTTTACCTGGAACACAGGC 977  
QY 2595 TGGGAAGTGTGAGAGCCAGAAACGAAACTGGAGAGCAGCTGAGAGATGACGCCACA 2654  
DB 976 TGGGAAGTGTGAGAGCCAGAAACGAAACTGGAGAGCAGCTGAGAGATGACGCCACA 917  
QY 2655 AAGCAGAGTGAAGAAATCGGCTGTGGAATGAGACAGATTCGGAGGCTCAGTCT 2714  
DB 916 AAGCAGAGTGAAGAAATCGGCTGTGGAATGAGACAGATTCGGAGGCTCAGTCT 857  
QY 2715 AAGCAGAGTGAAGAAATCGGAGTCAAGGCCAGCTCAGAGAGCTCAGGCTCTCCCT 2774  
DB 856 AAGCAGAGTGAAGAAATCGGAGTCAAGGCCAGCTCAGAGAGCTCAGGCTCTCCCT 797  
QY 2775 GCAGAGAGCGAGTCAAGTGTGACAGCCCTGAGAGCTGACGAGGCGGCTCGAGAGCCA 2834  
DB 796 GCAGAGAGCGAGTCAAGTGTGACAGCCCTGAGAGCTGACGAGGCGGCTCGAGAGCCA 737  
QY 2835 GCTTGGCCAGGCGAGAGACAGAGCTGGAAGAGCCACGAGAAAGCTGGAAGAGATCCA 2894  
DB 736 GCTTGGCCAGGCGAGAGACAGAGCTGGAAGAGCCACGAGAAAGCTGGAAGAGATCCA 677  
QY 2895 GGCATCCAGGCGAGTGAAGATGAATCCAGGCGAAATTTGATGCTCTTCTGTAACGCTG 2954  
DB 676 GGCATCCAGGCGAGTGAAGATGAATCCAGGCGAAATTTGATGCTCTTCTGTAACGCTG 617  
QY 2955 TACTGTAATCAAGACTGTGAGAGAGCACTAAACAGAGTGAACGAGAGCAAGCTGAACT 3014  
DB 616 TACTGTAATCAAGACTGTGAGAGAGCACTAAACAGAGTGAACGAGAGCAAGCTGAACT 557  
QY 3015 CAAACAACAAACTTCTACTGTTCTCAACCAACTGATGAGGCTTCTGGCGCAACGAGCA 3074



Db 556 CAACACGAAACTTCTACTTGTCCAAACACATCGATGAGGCTTCTGGCCGACGACGA 497  
 QY 3075 GATTGTACACTCGGAGTGAAGTGAACATCTCCGCCGAGATCAACGAGAGAGAT 3134  
 Db 496 GATTGTACACTCGGAGTGAAGTGAACATCTCCGCCGAGATCAACGAGAGAGAT 437  
 QY 3135 GCAGCTTCCAGCCGAGAGCAACGATGAGGCTCTGAAGACGACGATCATGCTGGA 3194  
 Db 436 GCAGCTTCCAGCCGAGAGCAACGATGAGGCTCTGAAGACGACGATCATGCTGGA 377  
 QY 3195 GGAACAGGTCAATGATTTGAGAGGCTTAAACGATGAGCTCTAGAAAAAGAGCGGAGTG 3254  
 Db 376 GGAACAGGTCAATGATTTGAGAGGCTTAAACGATGAGCTCTAGAAAAAGAGCGGAGTG 317  
 QY 3255 GGAAGCTTGAAGAGGCTCTCTGGTGAATGAGAAATCCAGTTTGAAGTGTGGGTTTGAAGA 3314  
 Db 316 GGAAGCTTGAAGAGGCTCTCTGGTGAATGAGAAATCCAGTTTGAAGTGTGGGTTTGAAGA 257  
 QY 3315 GCTGCAAGAGATGCTGACACCGAGAAACAGAGAGGCGGAGCCGATCAGCGGATCAC 3374  
 Db 256 GCTGCAAGAGATGCTGACACCGAGAAACAGAGAGGCGGAGCCGATCAGCGGATCAC 197  
 QY 3375 CGAGTCTGCCAGGTGTGTGAGCTGTGGCATGTAAGAGACAAAGCTTGAGATTCTGCTCT 3434  
 Db 196 CGAGTCTGCCAGGTGTGTGAGCTGTGGCATGTAAGAGACAAAGCTTGAGATTCTGCTCT 137  
 QY 3435 GCAGAGGCTCTCAAAAGACAGAAAGCTGAAGGCGGAGGCTCTCTGCAAGCTCAATGA 3494  
 Db 136 GCAGAGGCTCTCAAAAGACAGAAAGCTGAAGGCGGAGGCTCTCTGCAAGCTCAATGA 77  
 QY 3495 CCTGAGAGAAAGCATGCTATGCTTGAATGAATGCCCCGAAAGCTTACAGCAGAGCTGA 3554  
 Db 76 CCTGAGAG-----AGAGCTGGA 59  
 QY 3555 GACTGAAGAGAGCTCAAAAGAGGCTCTGAGAGAGCAAGCAATTAACGAGCGAG 3612  
 Db 58 GACTGAAGAGAGCTCAAAAGAGGCTCTGAGAGAGCAAGCAATTAACGAGCGAG 1  
 RESULT 15  
 ABZ68725  
 ID ABZ68725 standard; DNA; 1485 BP.  
 AC ABZ68725;  
 XX  
 DT 16-MAY-2003 (first entry)  
 XX  
 DE Nucleotide sequence of citron rho/rac-interacting kinase-short kinase.  
 KM Human; citron rho/rac-interacting kinase-short kinase; obesity;  
 KM chronic obstructive pulmonary disease; hypertension; diabetes;  
 KM coronary artery disease; hyperlipidaemia; stroke; gallbladder disease;  
 KM gout; osteoarthritis; sleep apnea; cancer; thrombotic disease;  
 KM polycystic ovarian syndrome; fertility; depression; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1485  
 FT /\*tag= a  
 FT /partial  
 FT /product= "citron rho/rac-interacting kinase-short  
 FT kinase"  
 XX  
 XX WO2003004629-A2.  
 XX  
 XX 16-JAN-2003.  
 XX  
 XX 01-JUL-2002; 2002MO-EP07229.  
 XX  
 XX 02-JUL-2001; 2001US-301853P.  
 XX 10-DEC-2001; 2001US-337130P.  
 XX 25-APR-2002; 2002US-375015P.  
 PR

XX (FAB ) BAYER AG.  
 PA  
 XX  
 XX Zhu Z;  
 PT  
 XX  
 DR WPI: 2003-221595/21.  
 DR P-P-SDB; ABP97681.  
 PT  
 XX  
 PT New human citron rho/rac-interacting kinase-short kinase polypeptide  
 PT and polynucleotide for preventing or treating diseases associated with  
 PT the polypeptide dysfunction, e.g. obesity or chronic obstructive  
 PT pulmonary disease  
 PS  
 PS Claim 1: Fig 1; 145pp; English.  
 CC The present sequence encodes a human citron rho/rac-interacting  
 CC kinase-short kinase polypeptide. The polynucleotide and polypeptide of  
 CC the invention are useful in preventing, ameliorating, or treating  
 CC diseases associated with the polypeptide dysfunction. The expression  
 CC vector or the reagent is useful in the preparation of a medicament for  
 CC modulating the activity of a human citron rho/rac-interacting  
 CC kinase-short kinase in a disease, such as obesity or chronic obstructive  
 CC pulmonary disease. These may also be used for treating obesity/  
 CC overweight-associated comorbidities, such as hypertension, diabetes,  
 CC coronary artery disease, hyperlipidaemia, stroke, gallbladder disease,  
 CC gout, osteoarthritis, sleep apnea, cancer, thrombotic diseases,  
 CC polycystic ovarian syndrome, reduced fertility, and depression.  
 XX  
 SQ Sequence 1485 BP; 373 A; 353 C; 372 G; 387 T; 0 other:  
 Query Match 22.7%; Score 1401.4; DB 25; Length 1485;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1402; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATTGTAAAGTTCAAAATATGAGAGCGGAAATCTTTGAGATGCTGCTGTAACCCATT 60  
 Db 1 ATTGTAAAGTTCAAAATATGAGAGCGGAAATCTTTGAGATGCTGCTGTAACCCATT 60  
 QY 61 GCCAGCCGGGCTCCAGGCTGAATCTGTTCTTCCAGGGGAAACACCTTTATGACTCAA 120  
 Db 61 GCCAGCCGGGCTCCAGGCTGAATCTGTTCTTCCAGGGGAAACACCTTTATGACTCAA 120  
 QY 121 CAGCAGATGCTCTCTTTTCCGAGAGGAGATTTAGATGCCCTCTTGTCTTTGAA 180  
 Db 121 CAGCAGATGCTCTCTTTTCCGAGAGGAGATTTAGATGCCCTCTTGTCTTTGAA 180  
 QY 121 CAGCAGATGCTCTCTTTTCCGAGAGGAGATTTAGATGCCCTCTTGTCTTTGAA 180  
 Db 121 CAGCAGATGCTCTCTTTTCCGAGAGGAGATTTAGATGCCCTCTTGTCTTTGAA 180  
 QY 181 GAATGAGTCAGGCTGCTCTGATGAAGATTAAAGCAGTGAACCTTGTCCGAGATAT 240  
 Db 181 GAATGAGTCAGGCTGCTCTGATGAAGATTAAAGCAGTGAACCTTGTCCGAGATAT 240  
 QY 241 TCCGACACCATAGCTAGTTAAGAGGCTCCAGCTTGGCAAGAGCTTGAAGTCA 300  
 Db 241 TCCGACACCATAGCTAGTTAAGAGGCTCCAGCTTGGCAAGAGCTTGAAGTCA 300  
 QY 241 TCCGACACCATAGCTAGTTAAGAGGCTCCAGCTTGGCAAGAGCTTGAAGTCA 300  
 Db 241 TCCGACACCATAGCTAGTTAAGAGGCTCCAGCTTGGCAAGAGCTTGAAGTCA 300  
 QY 301 AGTCTGTGGTGTGTGTGCTACCTTGTGAAGAGGAGGTGTAAAGAGAAAGCAACGG 360  
 Db 301 AGTCTGTGGTGTGTGTGCTACCTTGTGAAGAGGAGGTGTAAAGAGAAAGCAACGG 360  
 QY 301 AGTCTGTGGTGTGTGTGCTACCTTGTGAAGAGGAGGTGTAAAGAGAAAGCAACGG 360  
 Db 301 AGTCTGTGGTGTGTGTGCTACCTTGTGAAGAGGAGGTGTAAAGAGAAAGCAACGG 360  
 QY 361 GACATCTATGCTATGAAGATGAAGAGAGGCTTTATGAGCCAGAGAGAGTTCA 420  
 Db 361 GACATCTATGCTATGAAGATGAAGAGAGGCTTTATGAGCCAGAGAGAGTTCA 420  
 QY 361 GACATCTATGCTATGAAGATGAAGAGAGGCTTTATGAGCCAGAGAGAGTTCA 420  
 Db 361 GACATCTATGCTATGAAGATGAAGAGAGGCTTTATGAGCCAGAGAGAGTTCA 420  
 QY 421 TTTTGGAGGAGAGGAGGAGATTTATCTCCAGAGCAAGCCGAGATCCCAATTA 480  
 Db 421 TTTTGGAGGAGAGGAGGAGATTTATCTCCAGAGCAAGCCGAGATCCCAATTA 480  
 QY 481 CAGTATGCTTTGAGAGCAAAATCACTTTATCTGTCATGGAATATACGCTGAGGG 540  
 Db 481 CAGTATGCTTTGAGAGCAAAATCACTTTATCTGTCATGGAATATACGCTGAGGG 540  
 QY 541 GACTTGCTGCTCACTTTGAAATGATATGAGGACCAATTGATGAAACCTGATACGTT 600  
 Db 541 GACTTGCTGCTCACTTTGAAATGATATGAGGACCAATTGATGAAACCTGATACGTT 600



QY	601	TACCTAGCTGAGCTGATTTTGGCTTGCACAGCGCTTCATCTGATGGGATACGTGCATCGA	660
Db	601	TACCTAGCTGAGCTGATTTTGGCTTGCACAGCGCTTCATCTGATGGGATACGTGCATCGA	660
QY	661	GACATCAAGCCTGAGAAACATTTCTCGTGTGACCGCACAGACACATCAAGCTGGTGAATTTT	720
Db	661	GACATCAAGCCTGAGAAACATTTCTCGTGTGACCGCACAGACACATCAAGCTGGTGAATTTT	720
QY	721	GGATCTGCGCGGAAAATGAAATTCAACAAGTGGTGAATGCCAAATCCCGATTTGGAGCC	780
Db	721	GGATCTGCGCGGAAAATGAAATTCAACAAGTGGTGAATGCCAAATCCCGATTTGGAGCC	780
QY	781	CCAGATTACATGAGCTCCCTGGAAGTGTCTGATGTGAAACGGGGATGAAAAAGCACCTAC	840
Db	781	CCAGATTACATGAGCTCCCTGGAAGTGTCTGATGTGAAACGGGGATGAAAAAGCACCTAC	840
QY	841	GGCCTGGACTGTGACTGATGTGACTGATGGCGCTGATTTGCTATGATATTTATGGAGA	900
Db	841	GGCCTGGACTGTGACTGATGTGACTGATGGCGCTGATTTGCTATGATATTTATGGAGA	900
QY	901	TCCCCCTTCGAGAGGGAACTCTGCGCAGAACCTTCAATACATTTATGAAATTTCCACGG	960
Db	901	TCCCCCTTCGAGAGGGAACTCTGCGCAGAACCTTCAATACATTTATGAAATTTCCACGG	960
QY	961	TTTTTGAATTTCCAGATGACCCCAAGTGAAGCATTTCTTGAATCTGATTCAAAGC	1020
Db	961	TTTTTGAATTTCCAGATGACCCCAAGTGAAGCATTTCTTGAATCTGATTCAAAGC	1020
QY	1021	TTGTTGTGCGGCCGAAAAGAGAGACTGAAGTTTGAAGTCTTTTGCTGCCATCTTTCTTC	1080
Db	1021	TTGTTGTGCGGCCGAAAAGAGAGACTGAAGTTTGAAGTCTTTTGCTGCCATCTTTCTTC	1080
QY	1081	TCTTAAATTTGACTGGAACAACATTCGTATACCTCTCTCCCCCTTGTTCCACGCTCAAG	1140
Db	1081	TCTTAAATTTGACTGGAACAACATTCGTATACCTCTCTCCCCCTTGTTCCACGCTCAAG	1140
QY	1141	TCTGACGATGACACCTCCCAATTTTGATGAAACAGAGAAATTCGTGGGTTTCACTCT	1200
Db	1141	TCTGACGATGACACCTCCCAATTTTGATGAAACAGAGAAATTCGTGGGTTTCACTCT	1200
QY	1201	CCGTGCCAGCTGAGACCTCTCAGAGCTCTCCGGGTGAAGAACTGCCGTTTGTGGGGTTTTCG	1260
Db	1201	CCGTGCCAGCTGAGACCTCTCAGAGCTCTCCGGGTGAAGAACTGCCGTTTGTGGGGTTTTCG	1260
QY	1261	TACAGCAAGCACTGAGGATCTTTGGTGTGATCTGAGTGTGTTGTGCGGGCTGGAATCC	1320
Db	1261	TACAGCAAGCACTGAGGATCTTTGGTGTGATCTGAGTGTGTTGTGCGGGCTGGAATCC	1320
QY	1321	CTTGCAAGACTAGCTCCATGTAAGAAAGAACTTCTCATCAAAAGCAAAAGACTCAAGAC	1380
Db	1321	CTTGCAAGACTAGCTCCATGTAAGAAAGAACTTCTCATCAAAAGCAAAAGACTCAAGAC	1380
QY	1381	TCTCAGGCAAGTGTCAACAAGT 1403	
Db	1381	TCTCAGGCAAGTGTCAACAAGT 1403	

Search completed: November 15, 2003, 04:22:56  
Job time : 989.601 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 15, 2003, 01:08:15 ; Search time 14461.8 Seconds  
(without alignments)  
17439.597 Million cell updates/sec

Title: US-10-028-946-1

Perfect score: 6165  
Sequence: 1 atgtctgaagtcacataatg.....gggaccagctctcagratataa 6165

Scoring table: IDENTITY\_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenBank:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_ov:\*  
7: gb\_ov:\*  
8: gb\_ov:\*  
9: gb\_ov:\*  
10: gb\_ov:\*  
11: gb\_ov:\*  
12: gb\_ov:\*  
13: gb\_ov:\*  
14: gb\_ov:\*  
15: gb\_ov:\*  
16: gb\_ov:\*  
17: gb\_ov:\*  
18: gb\_ov:\*  
19: gb\_ov:\*  
20: gb\_ov:\*  
21: gb\_ov:\*  
22: gb\_ov:\*  
23: gb\_ov:\*  
24: gb\_ov:\*  
25: gb\_ov:\*  
26: gb\_ov:\*  
27: gb\_ov:\*  
28: gb\_ov:\*  
29: gb\_ov:\*  
30: gb\_ov:\*  
31: gb\_ov:\*  
32: gb\_ov:\*  
33: gb\_ov:\*  
34: gb\_ov:\*  
35: gb\_ov:\*  
36: gb\_ov:\*  
37: gb\_ov:\*  
38: gb\_ov:\*  
39: gb\_ov:\*  
40: gb\_ov:\*  
41: gb\_ov:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6165	100.0	6165	AX574425	AX574425 Sequence
2	6161.8	99.9	6298	AX504254	AX504254 Sequence
3	6160.2	99.9	6165	AX671105	AX671105 Sequence
4	6160.2	99.9	8603	AX671108	AX671108 Sequence
5	6142.2	99.6	6156	AX671112	AX671112 Sequence
6	6126	99.4	6159	AX166510	AX166510 Sequence
7	6121.8	99.3	6189	AX503780	AX503780 Sequence
8	6105.4	99.0	6201	AX503778	AX503778 Sequence
9	5843.2	94.8	8576	AX257469	AX257469 Homo sapi
10	5837	94.7	5877	AX574427	AX574427 Sequence
11	5661.4	91.8	6159	AX429514	AX429514 Sequence
12	5661.4	91.8	6574	AX429512	AX429512 Sequence
13	5105.2	82.8	6554	AF066824	AF066824 Mus muscu
14	4021.2	65.2	4967	AF070066	AF070066 Mus muscu
15	4018.2	65.2	5019	AF039218	AF039218 Rattus no
16	3906.6	63.4	5952	AF039218	AF039218 Rattus no
17	2818.2	45.7	5261	AF023166	AF023166 Homo sapi
18	2818.2	45.7	5261	AF023166	AF023166 Homo sapi
19	2403	39.0	3495	BC051165	BC051165 Mus muscu
20	2403	39.0	3506	BC023775	BC023775 Mus muscu
21	1401.4	22.7	1485	AX671037	AX671037 Sequence
22	1401.4	22.7	1785	AX671044	AX671044 Sequence
23	1399.8	22.7	2066	AX642956	AX642956 Sequence
24	1398.2	22.7	1515	AR253937	AR253937 Sequence
25	1088.8	17.7	2380	AX671042	AX671042 Sequence
26	1088.8	17.7	2380	AF066823	AF066823 Mus muscu
27	1080.8	17.5	1386	AF070065	AF070065 Rattus no
28	754	12.2	1058	AX053315	AX053315 Sequence
29	698	11.3	817	AX053416	AX053416 Sequence
30	350.4	5.7	354	BD059649	BD059649 Secreted
31	311.2	5.0	136436	AC002563	AC002563 Human PAC
32	311.2	5.0	147862	AC079317	AC079317 Homo sapi
33	311.2	5.0	190514	AC026363	AC026363 Homo sapi
34	311.2	5.0	208686	AC079262	AC079262 Homo sapi
35	258	4.2	258	AR269950	AR269950 Sequence
36	257.2	4.2	207404	AC115795	AC115795 Mus muscu
37	256.2	4.2	32658	AC135099	AC135099 Rattus no
38	256.2	4.2	234312	AC133614	AC133614 Rattus no
39	256.2	4.2	248557	AC097432	AC097432 Rattus no
40	250	4.1	254	BD071469	BD071469 Secreted
41	246.8	4.0	1530	AX504235	AX504235 Sequence
42	241.6	3.9	6335	AF021935	AF021935 Rattus no
43	235.2	3.8	4917	AY166586	AY166586 Homo sapi
44	233	3.8	2785	HS059305	HS059305 Human ser-t
45	233	3.8	4917	HSN518976	HSN518976 Homo sapi

# ALIGNMENTS

RESULT 1	AX574425	6165 bp	DNA	linear	PAT 07-JAN-2003
LOCUS	AX574425	Sequence 1 from Patent WO02059325.			
DEFINITION	AX574425				
ACCESSION	AX574425				
VERSION	AX574425.1	GI:27551751			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1	Yu, X., Miranda, M. and Fridde, C.J.			
AUTHORS	Human kinases and polynucleotides encoding the same				
TITLE	Human kinases and polynucleotides encoding the same				
JOURNAL	Patent: WO 02059325-A 1 01-AUG-2002;				

## Lexicon Genetics Incorporated (US)

FEATURES  
Location/Qualifierssource  
1. .6165  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 1735 a 1550 c 1679 g 1201 t  
ORIGINQuery Match 100.0%; Score 6165; DB 6; Length 6165;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTTAAAGTTCAAAATATGAGCGCGAATCCTTGGATGCTGCTGCTGTAACCAT 60  
DB 1 AGTTTAAAGTTCAAAATATGAGCGCGAATCCTTGGATGCTGCTGCTGTAACCAT 60  
QY 61 GCACGCGGCGCTCCAGCGCTGAATCTGTTCTCCAGGGGAAACCACTTTATGACTCA 120  
DB 61 GCACGCGGCGCTCCAGCGCTGAATCTGTTCTCCAGGGGAAACCACTTTATGACTCA 120  
QY 121 CAGCAGATGCTCTCTCTTCCGAGAGGAGATATAGATGCCCTCTTGTCTCTTGA 180  
DB 121 CAGCAGATGCTCTCTCTTCCGAGAGGAGATATAGATGCCCTCTTGTCTCTTGA 180  
QY 181 GAATGAGTCAGCTGCTCTGATGAAATTAAGCAGTGAATCTTGTCCGAAATAT 240  
DB 181 GAATGAGTCAGCTGCTCTGATGAAATTAAGCAGTGAATCTTGTCCGAAATAT 240  
QY 241 TCCGACACCATAGTGAAGTTCAGAGAGCTCCAGCCTTCGGCAAGAGACTTCGAAGTCGA 300  
DB 241 TCCGACACCATAGTGAAGTTCAGAGAGCTCCAGCCTTCGGCAAGAGACTTCGAAGTCGA 300  
QY 301 AGTCTTGAAGTGTGTGTCACTTTGCTGAAGTGAAGTGTGAAGAGAAACCAACCGG 360  
DB 301 AGTCTTGAAGTGTGTGTCACTTTGCTGAAGTGAAGTGTGAAGAGAAACCAACCGG 360  
QY 361 GACATCTATGCTATGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
DB 361 GACATCTATGCTATGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
QY 421 TTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
DB 421 TTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
QY 481 CAGATGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
DB 481 CAGATGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
QY 541 GACTTGTGCTCACTTTGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
DB 541 GACTTGTGCTCACTTTGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
QY 601 TACCTAGTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660  
DB 601 TACCTAGTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660  
QY 661 GACATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
DB 661 GACATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 721 GATCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
DB 721 GATCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
QY 781 CCGATATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
DB 781 CCGATATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
QY 841 GGGCTGAGCTGTGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 900  
DB 841 GGGCTGAGCTGTGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 900

QY 901 TCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 901 TCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 TTTTGAATTTTCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
DB 961 TTTTGAATTTTCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
QY 1021 TTGTTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
DB 1021 TTGTTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1081 TCTAAAATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
DB 1081 TCTAAAATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1141 TCTGAGATGAGACCTCCAAATTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
DB 1141 TCTGAGATGAGACCTCCAAATTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
QY 1201 CCGTGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
DB 1201 CCGTGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
QY 1261 TACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
DB 1261 TACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
QY 1321 CCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
DB 1321 CCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
QY 1381 TCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
DB 1381 TCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
QY 1441 GAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
DB 1441 GAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
QY 1501 TCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
DB 1501 TCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
QY 1561 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
DB 1561 TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
QY 1621 GATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
DB 1621 GATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
QY 1681 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
DB 1681 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
QY 1741 CCGAGTATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800  
DB 1741 CCGAGTATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800  
QY 1801 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
DB 1801 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
QY 1861 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
DB 1861 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
QY 1921 CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
DB 1921 CTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
QY 1981 AATATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040

```
Db 1981 AATATCCGCGAAGAGAGAGCGAGCGAGGGAGCTGGAGAACTGCGAAGACCGAGAG 2040
Qy 2041 GATTTCTTGAAGGCACTAGAAAAGAACTGGTGAAGCTGAGAACCGCCGCTTCTCTG 2100
Db 2041 GATTTCTTGAAGGCACTAGAAAAGAACTGGTGAAGCTGAGAACCGCCGCTTCTCTG 2100
Qy 2101 GAGAACAGAGTAAAGAGACTAGAGACATAGAGCGGTAGAGAAAACAGACTGAAAGATGAC 2160
Db 2101 GAGAACAGAGTAAAGAGACTAGAGACATAGAGCGGTAGAGAAAACAGACTGAAAGATGAC 2160
Qy 2161 ATCCAGCAAAAATCCCAACAGATCCAGAGATGAGCTGATAAATTTCTGAGCTGAAAGAG 2220
Db 2161 ATCCAGCAAAAATCCCAACAGATCCAGAGATGAGCTGATAAATTTCTGAGCTGAAAGAG 2220
Qy 2221 AAAACATCGGAGGGCCCAAGTCTCAGCCGACCACTAGAGAGTGCACCTGAAACAGAAAGAG 2280
Db 2221 AAAACATCGGAGGGCCCAAGTCTCAGCCGACCACTAGAGAGTGCACCTGAAACAGAAAGAG 2280
Qy 2281 CAGCACTATAGAGAAAAGATTAAAGTGTGACATCAATCAATPAAAGAAAGACCTGCTGAC 2340
Db 2281 CAGCACTATAGAGAAAAGATTAAAGTGTGACATCAATCAATPAAAGAAAGACCTGCTGAC 2340
Qy 2341 AAGGAGACACTGAGAGAAATAGATGACAGAGACAGAGAGAGGCGCCATGAGAGGGGAAA 2400
Db 2341 AAGGAGACACTGAGAGAAATAGATGACAGAGACAGAGAGAGGCGCCATGAGAGGGGAAA 2400
Qy 2401 ATTCTCAGCGAAAGAAAGGCGATGATCAATGCTATGATTCAGATTCAGATCCCTGAAA 2460
Db 2401 ATTCTCAGCGAAAGAAAGGCGATGATCAATGCTATGATTCAGATTCAGATCCCTGAAA 2460
Qy 2461 CAGAGGATTTGGAACTGTCTGAAAGCCATTAACCTTGACAGAAATAGAGCTTTTAAAC 2520
Db 2461 CAGAGGATTTGGAACTGTCTGAAAGCCATTAACCTTGACAGAAATAGAGCTTTTAAAC 2520
Qy 2521 CAAAGGAAATGAAAGGCGCAAGAGAGATGATTTCTGAACTCAGGCAAGAGAAATTTTAC 2580
Db 2521 CAAAGGAAATGAAAGGCGCAAGAGAGATGATTTCTGAACTCAGGCAAGAGAAATTTTAC 2580
Qy 2581 CTGAGAGACACAGGCTGGGAAATTGAGAGCCCAAGAACTGAGAGAGCAGCTGGAG 2640
Db 2581 CTGAGAGACACAGGCTGGGAAATTGAGAGCCCAAGAACTGAGAGAGCAGCTGGAG 2640
Qy 2641 AAGATCAGCCACCAAGACCAAGTGAACAAGATGAGGCTGGAACCTGAGAGACAAAGTTG 2700
Db 2641 AAGATCAGCCACCAAGACCAAGTGAACAAGATGAGGCTGGAACCTGAGAGACAAAGTTG 2700
Qy 2701 CGGAGAGTCACTTGAAGACAGAGAGCAGAGAACTGAGCTCAAGCCGACGCTCAAGAG 2760
Db 2701 CGGAGAGTCACTTGAAGACAGAGAGCAGAGAACTGAGCTCAAGCCGACGCTCAAGAG 2760
Qy 2761 CTACAGCTCTCCGCGAGAGCGGAGTCAAGTGAACGCTCGGAGAGCTGACAGGCGG 2820
Db 2761 CTACAGCTCTCCGCGAGAGCGGAGTCAAGTGAACGCTCGGAGAGCTGACAGGCGG 2820
Qy 2821 GCCCTGAGAGACAGCTTTCGCGAGAGCAGAGACAGAGCTGGAAGAGACCAAGAGAAAGCT 2880
Db 2821 GCCCTGAGAGACAGCTTTCGCGAGAGCAGAGACAGAGCTGGAAGAGACCAAGAGAAAGCT 2880
Qy 2881 GAAAGAGAGATCCAGGCACTTCAAGGCAATAGAGATGAATCCAGGCGCAATTTGATGCT 2940
Db 2881 GAAAGAGAGATCCAGGCACTTCAAGGCAATAGAGATGAATCCAGGCGCAATTTGATGCT 2940
Qy 2941 CTTGCTAACAGCTGTACTGTATACAGAGCCTGAGAGAGAGCTAAACAGCTGACGAG 3000
Db 2941 CTTGCTAACAGCTGTACTGTATACAGAGCCTGAGAGAGAGCTAAACAGCTGACGAG 3000
Qy 3001 GACACGCTGAATCTCAACCAACCAAACTTCTAATTCGCAACCACTCGATGAGGCTTCT 3060
Db 3001 GACACGCTGAATCTCAACCAACCAAACTTCTAATTCGCAACCACTCGATGAGGCTTCT 3060
Qy 3061 GGCGCCAGACGAGATTTGTACAACTGCAAGTGAAGTGAACATCTCCGCGGAGATC 3120
Db 3061 GGCGCCAGACGAGATTTGTACAACTGCAAGTGAAGTGAACATCTCCGCGGAGATC 3120
Qy 3121 ACGGAAAGAGATGCACTTACAGCCAGAAAGCAAAAGATGAGAGCTTGAAGACCAAG 3180
Db 3121 ACGGAAAGAGATGCACTTACAGCCAGAAAGCAAAAGATGAGAGCTTGAAGACCAAG 3180
Qy 3181 TGACACATGCTGAGAGAAAGAGTCAATGATTTTGAAGAGCCCTTAAAGATGAGCTGAGAA 3240
Db 3181 TGACACATGCTGAGAGAAAGAGTCAATGATTTTGAAGAGCCCTTAAAGATGAGCTGAGAA 3240
Qy 3241 AAAAGCGGCACTGAGAGGCGCTGAGAGAGCGCTCTGGTGAATGAATCCAGTTTGAAG 3300
Db 3241 AAAAGCGGCACTGAGAGGCGCTGAGAGAGCGCTCTGGTGAATGAATCCAGTTTGAAG 3300
Qy 3301 TGTCGGGCTTGAAGAGCTGAGAGAGATGCTGAGACACGAGAAAACAGAGAGAGAGGCG 3360
Db 3301 TGTCGGGCTTGAAGAGCTGAGAGAGATGCTGAGACACGAGAAAACAGAGAGAGAGGCG 3360
Qy 3361 GATCAGCGGATCAACGAGCTCTGCGCAGGCTGAGAGAGCTGAGAGAGAGAGAGAGAGGCT 3420
Db 3361 GATCAGCGGATCAACGAGCTCTGCGCAGGCTGAGAGAGCTGAGAGAGAGAGAGAGGCT 3420
Qy 3421 GAGATTTCTGCTCTGAGAGAGCTCTTCAAGAGCAGAACTGAGAGAGAGAGAGAGCTTCT 3480
Db 3421 GAGATTTCTGCTCTGAGAGAGCTCTTCAAGAGCAGAACTGAGAGAGAGAGAGAGCTTCT 3480
Qy 3481 GACAACTCAATGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTTA 3540
Db 3481 GACAACTCAATGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTTA 3540
Qy 3541 CAGCAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Db 3541 CAGCAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Qy 3601 TTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Db 3601 TTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Qy 3661 GAAAGCTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
Db 3661 GAAAGCTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
Qy 3721 GAAAACATTCAGAGTCTCTATTTCTCATGAAAAGGTGAAAAGTGAAGAGAGAGAGAGAG 3780
Db 3721 GAAAACATTCAGAGTCTCTATTTCTCATGAAAAGGTGAAAAGTGAAGAGAGAGAGAGAG 3780
Qy 3781 CAAACCAAACTCATGATTTTCTGCAAGCAGAAATGAGACCAACTGCTAAAAAGAGAAAG 3840
Db 3781 CAAACCAAACTCATGATTTTCTGCAAGCAGAAATGAGACCAACTGCTAAAAAGAGAAAG 3840
Qy 3841 GTTCTCTGCAATGATGATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Db 3841 GTTCTCTGCAATGATGATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Qy 3901 GAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Db 3901 GAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Qy 3961 GCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020
Db 3961 GCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020
Qy 4021 ATGCGCATGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Db 4021 ATGCGCATGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Qy 4081 GCGCCGCGATCCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Db 4081 GCGCCGCGATCCAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Qy 4141 AAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200
Db 4141 AAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200
```

QY	4201	ACAAAGTGTGCTGTGTGTCTTGGAATCCGTGGCACTTTGACGCGCAGAGCATCCAAATGTCTC	4266
Db	4201	ACAAAGTGTGCTGTGTGTCTTGGAATCCGTGGCACTTTGACGCGCAGAGCATCCAAATGTCTC	4266
QY	4261	GAATGTACAGTGAATGTGTCAACCCAGTGTCCACGTGTTCACGCCACTTGCCTGGCTTGG	4320
Db	4261	GAATGTACAGTGAATGTGTCAACCCAGTGTCCACGTGTTCACGCCACTTGCCTGGCTTGG	4320
QY	4321	CTGTGTAATATGTCGACACACTTTCACCGAGGCTTTCTGTGCGTGAACAATATGAATCTCCCA	4380
Db	4321	CTGTGTAATATGTCGACACACTTTCACCGAGGCTTTCTGTGCGTGAACAATATGAATCTCCCA	4380
QY	4381	GCTCTCCAGACCAAGAGGCCACAGCAGACCTTGACCTTGAAAGGTGATGAAGGTGCC	4440
Db	4381	GCTCTCCAGACCAAGAGGCCACAGCAGACCTTGACCTTGAAAGGTGATGAAGGTGCC	4440
QY	4441	AGGAATTAACAAACGAGGACAGCAAGGCTGGACAGAAAGTACATGTGCTTGAGAGATCA	4500
Db	4441	AGGAATTAACAAACGAGGACAGCAAGGCTGGACAGAAAGTACATGTGCTTGAGAGATCA	4500
QY	4501	AAAGTCTCATTTATATGACATGAAAGCCAGAGAGCTGGAACAGAGGCTGGTGAAGAAATTT	4560
Db	4501	AAAGTCTCATTTATGACATGAAAGCCAGAGAGCTGGAACAGAGGCTGGTGAAGAAATTT	4560
QY	4561	GAGCTGTGCTTCCGACCGGGATGTATCTATTCATGAGGCGTTGTGGTCTCCGAATC	4620
Db	4561	GAGCTGTGCTTCCGACCGGGATGTATCTATTCATGAGGCGTTGTGGTCTCCGAATC	4620
QY	4621	GCAATATACAGCCAAAGACAGATGTCCCATACATCTGAAGATGAAATCTCACCCGACACC	4680
Db	4621	GCAATATACAGCCAAAGACAGATGTCCCATACATCTGAAGATGAAATCTCACCCGACACC	4680
QY	4681	ACCTGTGGCCCGGGAGAACCCCTACTGTGCTAGTCCGACCTTCCCTGACAAACAGCCG	4740
Db	4681	ACCTGTGGCCCGGGAGAACCCCTACTGTGCTAGTCCGACCTTCCCTGACAAACAGCCG	4740
QY	4741	TGGGTACCCGCTTAGAAATCAGTTGTGACAGGTGGAGAGTTCTAGGAGAAAAGCAGAA	4800
Db	4741	TGGGTACCCGCTTAGAAATCAGTTGTGACAGGTGGAGAGTTCTAGGAGAAAAGCAGAA	4800
QY	4801	GCTGATGCTAAACCTGCTTGAAACCTCCTGCTGAACTGGAAGGTGATGACCGCTTAGAC	4860
Db	4801	GCTGATGCTAAACCTGCTTGAAACCTCCTGCTGAACTGGAAGGTGATGACCGCTTAGAC	4860
QY	4861	ATGAATCAGACGCTCCCTCAGTGACACAGGTGGTGGTGGGGCACGAGAAAGGCTC	4920
Db	4861	ATGAATCAGACGCTCCCTCAGTGACACAGGTGGTGGTGGGGCACGAGAAAGGCTC	4920
QY	4921	TACGCTCTGAAATGTCTTGAAAACTCCCTAACCCATGTCCAGGAATTTGAGACATCTTC	4980
Db	4921	TACGCTCTGAAATGTCTTGAAAACTCCCTAACCCATGTCCAGGAATTTGAGACATCTTC	4980
QY	4981	CAAAATTTATTTATCAAGACCTTGAGAGTATCATGATAGGAGAGAAAGGCGAGCA	5040
Db	4981	CAAAATTTATTTATCAAGACCTTGAGAGTATCATGATAGGAGAGAAAGGCGAGCA	5040
QY	5041	CTGTGTCTTGAGACGTGAAGAAAGTAAAACAGTCCCTGGCCCAAGTCCCACTGCTGCC	5100
Db	5041	CTGTGTCTTGAGACGTGAAGAAAGTAAAACAGTCCCTGGCCCAAGTCCCACTGCTGCC	5100
QY	5101	CAGCCCGACATCTCACCCCAACATTTTGAAGCTGTCAAGGCTGGCACTTTGTTGGGGCA	5160
Db	5101	CAGCCCGACATCTCACCCCAACATTTTGAAGCTGTCAAGGCTGGCACTTTGTTGGGGCA	5160
QY	5161	GGCAAGATTGAAACGGGCTCTGCACTGTGACAGCATGCGCAGCAAGTGTGCTATCTC	5220
Db	5161	GGCAAGATTGAAACGGGCTCTGCACTGTGACAGCATGCGCAGCAAGTGTGCTATCTC	5220
QY	5221	CGCTACACAGAAACCTCAGCAAAATATGTGATCCGAGAAAGATAGACTCTGAGAGCC	5280
Db	5221	CGCTACACAGAAACCTCAGCAAAATATGTGATCCGAGAAAGATAGACTCTGAGAGCC	5280

QY	5281	TCGACCTGATATCACCTTACACCAATTAAGTATCCATTGGAGCAATAAATTCTAGAA	5341
Dp	5281	TGCAGCTGATATCACCTTACACCAATTAAGTATCCATTGGAGCAATAAATTCTAGAA	5341
QY	5341	ATCGACATGAAAGCAGTATACACGCTCGAGGAATTCGTATTAAGATGACCATTCCTTGCA	5400
Dp	5341	ATCGACATGAAAGCAGTATACACGCTCGAGGAATTCGTATTAAGATGACCATTCCTTGCA	5400
QY	5401	CCTGCTGTGTTTGGCCGCTCTTCCACAGCTTCCCTGTCTCAATGTGCAAGTGAACAGC	5461
Dp	5401	CCTGCTGTGTTTGGCCGCTCTTCCACAGCTTCCCTGTCTCAATGTGCAAGTGAACAGC	5461
QY	5461	GCAGGGCAGCGAGAGAGTACTTGCTGTGTTTCCACGAATTTGAGAGTTGTGGATCTC	5521
Dp	5461	GCAGGGCAGCGAGAGAGTACTTGCTGTGTTTCCACGAATTTGAGAGTTGTGGATCTC	5521
QY	5521	TACGGAAAGCCTGATCGCCACAGACGATCTCAAGTGAAGTCCGTTACCTTTGGCCTTTGCC	5581
Dp	5521	TACGGAAAGCCTGATCGCCACAGACGATCTCAAGTGAAGTCCGTTACCTTTGGCCTTTGCC	5581
QY	5581	TACAGAGAAACCTATCTGTTTGTGACCCACTTCAACTCACTCGAAGTAAATTGAAATCCAG	5641
Dp	5581	TACAGAGAAACCTATCTGTTTGTGACCCACTTCAACTCACTCGAAGTAAATTGAAATCCAG	5641
QY	5641	GCAGGCTCTCAGAGAGGAGCCCTCGCCGAGCGTACTTGACATCCCGAACCCGCGCTAC	5701
Dp	5641	GCAGGCTCTCAGAGAGGAGCCCTCGCCGAGCGTACTTGACATCCCGAACCCGCGCTAC	5701
QY	5701	CTGGGCCCCCTGCATTTCTCTCAAGAGCGATTTACTTGGCGCTCTATACAGAGATTAATTA	5761
Dp	5701	CTGGGCCCCCTGCATTTCTCTCAAGAGCGATTTACTTGGCGCTCTATACAGAGATTAATTA	5761
QY	5761	AGGGTCAATTTGCTCCAGAGGGAACCTGTGAAGAGTCCGGCACTGAACACCAACGGGAGC	5821
Dp	5761	AGGGTCAATTTGCTCCAGAGGGAACCTGTGAAGAGTCCGGCACTGAACACCAACGGGAGC	5821
QY	5821	CCGTTCACCTTCCCGCAGCAGCCCCAATAAGAGAGCCCACTTCAACAGAGCAATC	5881
Dp	5821	CCGTTCACCTTCCCGCAGCAGCCCCAATAAGAGAGCCCACTTCAACAGAGCAATC	5881
QY	5881	ACCAAGCGCTGAGCTCCAGAGCCGACCCGAGAGGCCCCGAGCCACCCGCGAGAGCA	5941
Dp	5881	ACCAAGCGCTGAGCTCCAGAGCCGACCCGAGAGGCCCCGAGCCACCCGCGAGAGCA	5941
QY	5941	AGCAACATCCCAACCGCTACCGCGAGGAGCGAGCCGAGCTGCGCAGAGGAACAACTCTCGGC	6001
Dp	5941	AGCAACATCCCAACCGCTACCGCGAGGAGCGAGCCGAGCTGCGCAGAGGAACAACTCTCGGC	6001
QY	6001	CGCCCCCTGAGAGGAGAGAGTCCCCCGGCGGATATCTACGACCGCGAGAGAGCGGTC	6061
Dp	6001	CGCCCCCTGAGAGGAGAGAGTCCCCCGGCGGATATCTACGACCGCGAGAGAGCGGTC	6061
QY	6061	CCCGCGAGGCTGTTTGAAGACAGACAGAGGAGCCGAGCTGCTGCGGAGAGCGTGAAGAC	6121
Dp	6061	CCCGCGAGGCTGTTTGAAGACAGACAGAGGAGCCGAGCTGCTGCGGAGAGCGTGAAGAC	6121
QY	6121	CCCGCTGCCAAGGTGAACAAAGTCTTGAGACCACTTTCAGATTA 6165	
Dp	6121	CCCGCTGCCAAGGTGAACAAAGTCTTGAGACCACTTTCAGATTA 6165	

RESULT 2					
AX504254					
LOCUS	AX504254	6298 bp	DNA	linear	PAT 27-SEP-20
DEFINITION	Sequence 43 from Patent WO0233099.				
ACCESSION	AX504254				
VERSION	AX504254.1	GI:2386094			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
	Homo sapiens				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
REFERENCE	1				

## AUTHORS

Gururajan, R., Baughn, M. R., Wajia, N. K., Elliott, V. S., Xu, Y.,  
Aryizu, C., Yao, M. G., Ramkumar, J., Ding, L., Tang, Y. T., Hafila, A. J.,  
Nguyen, D. B., Gandhi, A. R., Lu, Y., Yue, H., Burford, N., Bandman, O.,  
Tribouley, C. M., Lal, P. G., Recipon, S. A., Lu, D. A., Borowsky, M. L.,  
Thornston, M., Swartner, A., Nangavelu, K., Khan, F. A. and Ison, C. H.

Human kinases  
Patent: WO 0233099-A 43 25-APR-2002;

Incyte Genomics, Inc. (US)  
Location/Qualifiers  
1. 6298

## FEATURES

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 7484498CB1"

BASE COUNT 1772 a 1585 c 1720 g 1221 t

## ORIGIN

Query Match 99.9%; Score 6161.8; DB 6; Length 6298;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6163; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 ATGTTGAAGTCAATATGAGCGCGAATCCTTTGAGTGTGCTGCTGAACCAT 60  
55 ATGTTGAAGTCAATATGAGCGCGAATCCTTTGAGTGTGCTGCTGAACCAT 114  
61 GCCGCGCGGCTCCAGGCTGAATCTGTTCTTCCAGGGAAACCACTTTAGTCAA 120  
115 GCCGCGCGGCTCCAGGCTGAATCTGTTCTTCCAGGGAAACCACTTTAGTCAA 174  
121 CAGGAGATGTCCTCTTCCCGAGAGGATATAGATGCCCTTGTCTCTTGA 180  
175 CAGGAGATGTCCTCTTCCCGAGAGGATATAGATGCCCTTGTCTCTTGA 234  
181 GAATGAGTCAAGCTGCTGTGATGAAGATTAAAGCAGTGAACAATTGTCCGGAATAT 240  
235 GAATGAGTCAAGCTGCTGTGATGAAGATTAAAGCAGTGAACAATTGTCCGGAATAT 294  
241 TCCGACACCAATGAGTGAATACAGGAGCTCCAGCTCCGCAAGAGACTCCAGAGTCA 300  
295 TCCGACACCAATGAGTGAATACAGGAGCTCCAGCTCCGCAAGAGACTCCAGAGTCA 354  
301 AGCTTGTAGGTTGTGCTCACTTGTCTGAAGTGAAGTGAAGAGAAAGCAACGGG 360  
355 AGCTTGTAGGTTGTGCTCACTTGTCTGAAGTGAAGTGAAGAGAAAGCAACGGG 414  
361 GACATCTATGCTATGAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
415 GACATCTATGCTATGAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 474  
421 TTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
475 TTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 534  
481 CAGTATGCTCTTCAAGCAAAATCACTTTATCTGCTCATGGAATATCAAGCTGAGAG 540  
535 CAGTATGCTCTTCAAGCAAAATCACTTTATCTGCTCATGGAATATCAAGCTGAGAG 594  
541 GACTTGTGCTCACTTTGAATAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
595 GACTTGTGCTCACTTTGAATAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 654  
601 TACTAGCTGAGCTGATTTTGGCTGTTCAAGGCTTCACTGATGGAATATCAAGCTGAG 660  
655 TACTAGCTGAGCTGATTTTGGCTGTTCAAGGCTTCACTGATGGAATATCAAGCTGAG 714  
661 GACATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
715 GACATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774  
721 GGAATCTGCGCGAAGATGAATCAACCAAGATGATGAATGCAATCCGATTTGGAGCC 780  
775 GGAATCTGCGCGAAGATGAATCAACCAAGATGATGAATGCAATCCGATTTGGAGCC 834

781 CCAGATTAATGAGCTCTGAAGTGTGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
835 CCAAGTTTACATGCTCTGAGTGTGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 894  
841 GGCCTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
895 GGCCTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 954  
901 TCCGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
955 TCCGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1014  
961 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
1015 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1074  
1021 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
1075 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1134  
1081 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
1135 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1194  
1141 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
1195 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1254  
1201 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
1255 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1314  
1261 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
1315 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374  
1321 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
1375 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434  
1381 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
1435 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494  
1441 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
1495 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554  
1501 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
1555 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614  
1561 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
1615 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674  
1621 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1675 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734  
1681 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
1735 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794  
1741 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800  
1795 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854  
1801 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
1855 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914  
1861 TTTTGAATTTCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920



```
Db 1915 GTGGGAAATATGCGAACTGGAGGATCAATGCTGAGACAGACTCAAAATTCAGAG 1974
Qy 1921 CTCGAAGAGAACTGGAGAGGCTGTAAAGCCAGACGAGGACCCAGCTGCTGAG 1980
Db 1975 CTCGAAGAGAACTGGAGAGGCTGTAAAGCCAGACGAGGACCCAGCTGCTGAG 2034
Qy 1981 AATATCCCGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
Db 2035 AATATCCCGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2094
Qy 2041 GATTCCTTGAAGAGCATCGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100
Db 2095 GATTCCTTGAAGAGCATCGAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2154
Qy 2101 GAGAGCAAGGTAAAGAGACTGAGAGCATGAGAGCTGAGAGAGAGAGAGAGAGAG 2160
Db 2155 GAGAGCAAGGTAAAGAGACTGAGAGCATGAGAGCTGAGAGAGAGAGAGAGAGAG 2214
Qy 2161 ATCCAGACAAATATCCAGACAGATCCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
Db 2215 ATCCAGACAAATATCCAGACAGATCCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2274
Qy 2221 AAACATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
Db 2275 AAACATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2334
Qy 2281 CAGCACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Db 2335 CAGCACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2394
Qy 2341 AAGAGAGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db 2395 AAGAGAGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2454
Qy 2401 ATTCTCAGCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
Db 2455 ATTCTCAGCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2514
Qy 2461 CAGAGAGTTGTGAACTGTCTGAAGCCAAATTAACCTGAGCAAAATGAGCATCTTTTACC 2520
Db 2515 CAGAGAGTTGTGAACTGTCTGAAGCCAAATTAACCTGAGCAAAATGAGCATCTTTTACC 2574
Qy 2521 CAAAGAGAACTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2580
Db 2575 CAAAGAGAACTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2634
Qy 2581 CTGAGAGCAAGGCTGGAGAGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2640
Db 2635 CTGAGAGCAAGGCTGGAGAGTTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2694
Qy 2641 AAGATCAGCCAGCAAGAGCAAGTGAACAAGATCGGCTGTGAGAACTGAGAGCAAGATTG 2700
Db 2695 AAGATCAGCCAGCAAGAGCAAGTGAACAAGATCGGCTGTGAGAACTGAGAGCAAGATTG 2754
Qy 2701 CCGGAGGTTCACTTAAGAGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2760
Db 2755 CCGGAGGTTCACTTAAGAGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2814
Qy 2761 CTACAGCTCTCCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2820
Db 2815 CTACAGCTCTCCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2874
Qy 2821 GCCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2880
Db 2875 GCCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2934
Qy 2881 GAAGAGAGATCCAGAGCACTCAAGGAGCATAGAGATGAATTCAGGAGCAAAATTTGATGCT 2940
Db 2935 GAAGAGAGATCCAGAGCACTCAAGGAGCATAGAGATGAATTCAGGAGCAAAATTTGATGCT 2994
Qy 2941 CTTCGTAACAGCTGTATCTGTAATCAAGAGCTTGAGAGAGAGGAGGAGGAGGAGGAG 3000
```

---

```
Db 2995 CTTCGTAACAGCTGTATCTGTAATCAAGAGCTTGAGAGAGGAGGAGGAGGAGGAGGAG 3054
Qy 3001 GACAACGCTGATCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 3060
Db 3055 GACAACGCTGATCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 3114
Qy 3061 GGGCCCAACGAGAGATTTGATCAACTGCGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3120
Db 3115 GGGCCCAACGAGAGATTTGATCAACTGCGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3174
Qy 3121 ACGGAGAGAGATGAGAGCTTACAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
Db 3175 ACGGAGAGAGATGAGAGCTTACAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3234
Qy 3181 TGACCAATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Db 3235 TGACCAATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3294
Qy 3241 AAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3300
Db 3295 AAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3354
Qy 3301 TGTGGGTTTCAAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
Db 3355 TGTGGGTTTCAAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3414
Qy 3361 GATCAGCGGATCAACGAGCTCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3420
Db 3415 GATCAGCGGATCAACGAGCTCTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3474
Qy 3421 GAGATTCCTGCTGAGAGAGGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
Db 3475 GAGATTCCTGCTGAGAGAGGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3534
Qy 3481 GACAAAGCTCAATGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
Db 3535 GACAAAGCTCAATGAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3594
Qy 3541 CAGAGAGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Db 3595 CAGAGAGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3654
Qy 3601 TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Db 3655 TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3714
Qy 3661 GAGGCTAGAGTTGGGCTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
Db 3715 GAGGCTAGAGTTGGGCTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3774
Qy 3721 GAAAAATTCAGAGTTCTTAATTCATGAAAGGTGAAGATGAAGGAGCAATTTCTCAA 3780
Db 3775 GAAAAATTCAGAGTTCTTAATTCATGAAAGGTGAAGATGAAGGAGCAATTTCTCAA 3834
Qy 3781 CAAACCAAACTCATTTGATTTTCTGCAAGCCAAATGAGCAACCACTGCTAAAGAAAGAAAG 3840
Db 3835 CAAACCAAACTCATTTGATTTTCTGCAAGCCAAATGAGCAACCACTGCTAAAGAAAGAAAG 3894
Qy 3841 GTTCCTCTGAGTCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3900
Db 3895 GTTCCTCTGAGTCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3954
Qy 3901 GAGCTAGAGAGAGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Db 3955 GAGCTAGAGAGAGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4014
Qy 3961 GCCCAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020
Db 4015 GCCCAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4074
Qy 4021 ATGCGCATGTCCGCAATGCTGCGGTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Db 4075 ATGCGCATGTCCGCAATGCTGCGGTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4134
```





Db 1801 CGAAAAGCAGCATGTCTGATTAACCTGTGAGGCTAAGATCAAGGGAAGCTGAA 1860  
Qy 1861 GTGGAGAAATATGCGAAACTGGAAGATCAATCTGACAGAGCTCAAAATTCGAGAG 1920  
Db 1861 GTGGAGAAATATGCGAAACTGGAAGATCAATCTGACAGAGCTCAAAATTCGAGAG 1920  
Qy 1921 CTCGAGAGAACTGGAAGAGCTGTAAAAAGCCAGCAGAGGCTCAAGAGCTGTGAG 1980  
Db 1921 CTCGAGAGAACTGGAAGAGCTGTAAAAAGCCAGCAGAGGCTCAAGAGCTGTGAG 1980  
Qy 1981 AATTTCGCGCAAGGCAAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040  
Db 1981 AATTTCGCGCAAGGCAAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040  
Qy 2041 GATTCCTCTGAGAGCATCAGAAAGAGCTGTGAGAGCTGAGAGAGCGCGGCTGTCG 2100  
Db 2041 GATTCCTCTGAGAGCATCAGAAAGAGCTGTGAGAGCTGAGAGAGCGCGGCTGTCG 2100  
Qy 2101 GAGAACAGAGTAAAGAGACTAGAGACCATGAGGCTGAGAGAAACAGACTGAAGATGAC 2160  
Db 2101 GAGAACAGAGTAAAGAGACTAGAGACCATGAGGCTGAGAGAAACAGACTGAAGATGAC 2160  
Qy 2161 ATCCAGCAAAATCCCAACAGATCCAGAGATGGCTGATTAATTCGAGGCTCGAAGAG 2220  
Db 2161 ATCCAGCAAAATCCCAACAGATCCAGAGATGGCTGATTAATTCGAGGCTCGAAGAG 2220  
Qy 2221 AAACATCGGAGGCGCCAAAGTCTGAGCCGAGACCTGAGAGCTGAGAAAGAGAG 2280  
Db 2221 AAACATCGGAGGCGCCAAAGTCTGAGCCGAGACCTGAGAGCTGAGAAAGAGAG 2280  
Qy 2281 CAGACATATGAGGAAAGATTTAAAGTGTGAGCAATCAAGTAAAGAAAGAGCTGTGAC 2340  
Db 2281 CAGACATATGAGGAAAGATTTAAAGTGTGAGCAATCAAGTAAAGAAAGAGCTGTGAC 2340  
Qy 2341 AAGGAGCACTGAGAGACATGATGACAGAGACAGAGAGAGGCGCATGAGAGGCAAA 2400  
Db 2341 AAGGAGCACTGAGAGACATGATGACAGAGACAGAGAGAGGCGCATGAGAGGCAAA 2400  
Qy 2401 ATTCTCAGCGAACAAGAGGCGATGATTAATGCTATGATTCGAAAGTTCGATTCCTGAG 2460  
Db 2401 ATTCTCAGCGAACAAGAGGCGATGATTAATGCTATGATTCGAAAGTTCGATTCCTGAG 2460  
Qy 2461 CAGAGATTTGAGAGCTGTCTGAGAGCCAAATTAATTCGAGCAAAATGAGCTTTTAC 2520  
Db 2461 CAGAGATTTGAGAGCTGTCTGAGAGCCAAATTAATTCGAGCAAAATGAGCTTTTAC 2520  
Qy 2521 CAAAGGAACTGAAAGGCGCAAGAGAGATGATTTCTGAACTCAGGCAACAGAAATTTAC 2580  
Db 2521 CAAAGGAACTGAAAGGCGCAAGAGAGATGATTTCTGAACTCAGGCAACAGAAATTTAC 2580  
Qy 2581 CTGGAGACAGAGCTGGAGAGTTGAGAGGCGCAAGAACGAGAACTGAGAGAGAGCTGAG 2640  
Db 2581 CTGGAGACAGAGCTGGAGAGTTGAGAGGCGCAAGAACGAGAACTGAGAGAGAGCTGAG 2640  
Qy 2641 AAGATCGCCACCAAGACCAAGATGACAGAAATCGGCTGTGAGACTGAGAGCAAAATTG 2700  
Db 2641 AAGATCGCCACCAAGACCAAGATGACAGAAATCGGCTGTGAGACTGAGAGCAAAATTG 2700  
Qy 2701 CGGAGAGTCACTGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760  
Db 2701 CGGAGAGTCACTGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760  
Qy 2761 CTACAGCTCTCCTGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820  
Db 2761 CTACAGCTCTCCTGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820  
Qy 2821 GCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880  
Db 2821 GCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880  
Qy 2881 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940  
Db 2881 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940

Qy 2941 CTTGCTAACAGCTGATCTGATATCAAGACCTGAGAGAGAGAGAGAGAGAGAGAG 3000  
Db 2941 CTTGCTAACAGCTGATCTGATATCAAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 3000  
Qy 3001 GACAAAGCTGAACTCAACCAACCAAACTTCTACTTGTCCAAACAACTGATGAGGCTCT 3060  
Db 3001 GACAAAGCTGAACTCAACCAACCAAACTTCTACTTGTCCAAACAACTGATGAGGCTCT 3060  
Qy 3061 GGGCGCAACGAGAGATGATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120  
Db 3061 GGGCGCAACGAGAGATGATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120  
Qy 3121 ACGGAAAGAGATGACAGCTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
Db 3121 ACGGAAAGAGATGACAGCTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
Qy 3181 TGACCATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240  
Db 3181 TGACCATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240  
Qy 3241 AAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300  
Db 3241 AAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300  
Qy 3301 TGTGCGGTTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360  
Db 3301 TGTGCGGTTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360  
Qy 3361 GATCAGCGGATCAACGAGTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420  
Db 3361 GATCAGCGGATCAACGAGTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420  
Qy 3421 GAGATTTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480  
Db 3421 GAGATTTCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480  
Qy 3481 GACAAAGCTCAATGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
Db 3481 GACAAAGCTCAATGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
Qy 3541 CAGCAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600  
Db 3541 CAGCAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600  
Qy 3601 TTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
Db 3601 TTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
Qy 3661 GAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720  
Db 3661 GAGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720  
Qy 3721 GAAAAATTCAGAGTCTCTATTTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780  
Db 3721 GAAAAATTCAGAGTCTCTATTTCTATGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780  
Qy 3781 CAAACCAAACTCATTTGATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840  
Db 3781 CAAACCAAACTCATTTGATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840  
Qy 3841 GTTCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900  
Db 3841 GTTCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900  
Qy 3901 GAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960  
Db 3901 GAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960  
Qy 3961 GCGCAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020  
Db 3961 GCGCAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020

QY 4021 ATGCGCATGTCGCGCATGCGGTGCGGAGACACAGCCAGTGCATGAGCCTGCTG 4080  
 Db 4021 ATGCGCATGTCGCGCATGCGGTGCGGAGACACAGCCAGTGCATGAGCCTGCTG 4080  
 QY 4081 GCCCGCCCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGAAATTTAGTGGGCTTT 4140  
 Db 4081 GCCCGCCCATCCAGCCGAGAAAGAGTCTTCACTCCAGAGAAATTTAGTGGGCTTT 4140  
 QY 4141 AAGGAAACGATGACACACATATTTCTCAACGATTTAAAGTAGACTGAACATGAGGCC 4200  
 Db 4141 AAGGAAACGATGACACACATATTTCTCAACGATTTAAAGTAGACTGAACATGAGGCC 4200  
 QY 4201 ACAAGTGTCTGTGTCTGATACCGTACCTTGACCTTGAGAGCCAGGACCTCAATGTCTC 4260  
 Db 4201 ACAAGTGTCTGTGTCTGATACCGTACCTTGAGAGCCAGGACCTCAATGTCTC 4260  
 QY 4261 GAATGTGAGTGTGTGTCACTCCAGAGTCTCAAGTCTTGCACGACCTTGGCGCTTG 4320  
 Db 4261 GAATGTGAGTGTGTGTCACTCCAGAGTCTCAAGTCTTGCACGACCTTGGCGCTTG 4320  
 QY 4321 CCGTGAATATGCGCACACCTTCAAGGAGCTTGTGCGGTGACAAATGAACTGCCCA 4380  
 Db 4321 CCGTGAATATGCGCACACCTTCAAGGAGCTTGTGCGGTGACAAATGAACTGCCCA 4380  
 QY 4381 GATCTCCAGACCAAGGAGCCAGACAGCTTGCACCTTGAAAGATGAAAGTGGCC 4440  
 Db 4381 GATCTCCAGACCAAGGAGCCAGACAGCTTGCACCTTGAAAGATGAAAGTGGCC 4440  
 QY 4441 AGGAATATGAAAGAGGACGCAAGGCTGGGACAGAGATGACATTTCTCTGAGAGGATCA 4500  
 Db 4441 AGGAATATGAAAGAGGACGCAAGGCTGGGACAGAGATGACATTTCTCTGAGAGGATCA 4500  
 QY 4501 AAGTCTCATTTATGACAAATGAAAGCAGAGAGTGTGACAGAGCCGCTGAGAAATTT 4560  
 Db 4501 AAGTCTCATTTATGACAAATGAAAGCAGAGAGTGTGACAGAGCCGCTGAGAAATTT 4560  
 QY 4561 GAGTGTGCTTCCCGAGGAGATGATATCTATCTAGTGTGCTGTGTCTTCCGAACTC 4620  
 Db 4561 GAGTGTGCTTCCCGAGGAGATGATATCTATCTAGTGTGCTGTGTCTTCCGAACTC 4620  
 QY 4621 GCAAAATACAGCAAAAGAGATGTCCCATATCATATCTGAAATCTGACCCGACAC 4680  
 Db 4621 GCAAAATACAGCAAAAGAGATGTCCCATATCATATCTGAAATCTGACCCGACAC 4680  
 QY 4681 ACCGTGTGCTTCCCGAGGAGATGATATCTATCTAGTGTGCTGTGTCTTCCGAACTC 4740  
 Db 4681 ACCGTGTGCTTCCCGAGGAGATGATATCTATCTAGTGTGCTGTGTCTTCCGAACTC 4740  
 QY 4741 TGGGTACCGCTTGAATCAGTTGTGACAGTGTGGAGAGTTTCTAGGAGAAAGAGAA 4800  
 Db 4741 TGGGTACCGCTTGAATCAGTTGTGACAGTGTGGAGAGTTTCTAGGAGAAAGAGAA 4800  
 QY 4801 GCTGATCTTAACTGCTTGAATACTCTCTGTAATCTGAAAGTATGACCTCTTAAAGC 4860  
 Db 4801 GCTGATCTTAACTGCTTGAATACTCTCTGTAATCTGAAAGTATGACCTCTTAAAGC 4860  
 QY 4861 ATGAATCTGACGCGCTTCAAGTACAGAGTGTGTGCTGAGGAGGAGGAGGCTC 4920  
 Db 4861 ATGAATCTGACGCGCTTCAAGTACAGAGTGTGTGCTGAGGAGGAGGAGGCTC 4920  
 QY 4921 TAGCGCTGATGTCTTGAATACTCTCTTAACTGTAATCTGAAAGTATGAGCAGTCTT 4980  
 Db 4921 TAGCGCTGATGTCTTGAATACTCTCTTAACTGTAATCTGAAAGTATGAGCAGTCTT 4980  
 QY 4981 CAAATTTATATTAATCAAGGAGCTGAGAGGAGTCTGATATGACAGAGAGAGAGGAGG 5040  
 Db 4981 CAAATTTATATTAATCAAGGAGCTGAGAGGAGTCTGATATGACAGAGAGAGAGGAGG 5040  
 QY 5041 CTGTGTCTTGTGAGAGTGAAGAAAGTGAACAGTCCCTGAGCCAGTCCCACTGCTGCTC 5100  
 Db 5041 CTGTGTCTTGTGAGAGTGAAGAAAGTGAACAGTCCCTGAGCCAGTCCCACTGCTGCTC 5100  
 QY 5101 CAGCGGACATCTACCCCAATTTTGAAGCTGTCAAGGCTGCCACTTGTGTGGGCA 5160

Db 5101 CAGCGGACATCTACCCCAATTTTGAAGCTGTCAAGGCTGCCACTTGTGTGGGCA 5160  
 QY 5161 GCGAAGATTTGAGAGCGGCTCTGATCTGTGACGCAATGCCAGAAAGTGTATCTC 5220  
 Db 5161 GCGAAGATTTGAGAGCGGCTCTGATCTGTGACGCAATGCCAGAAAGTGTATCTC 5220  
 QY 5221 CGCTACACGAAACCTCAGCAATACTGATCCGGAAGAGATGAGACTCAGAGCC 5280  
 Db 5221 CGCTACACGAAACCTCAGCAATACTGATCCGGAAGAGATGAGACTCAGAGCC 5280  
 QY 5281 TGCAGTGTATCTACCTTCAATTAAGATCTTCAATTTGAAACCAATTAATTTAGAA 5340  
 Db 5281 TGCAGTGTATCTACCTTCAATTAAGATCTTCAATTTGAAACCAATTAATTTAGAA 5340  
 QY 5341 ATGACATGAAAGAGTACAGCTTGAAGATTTCTGATTAAGATGACATTTCTTGGCA 5400  
 Db 5341 ATGACATGAAAGAGTACAGCTTGAAGATTTCTGATTAAGATGACATTTCTTGGCA 5400  
 QY 5401 CCGTGTGTGTGCGGCTCTTCAAGAGCTTCCCTGTCTCAATCTGAGAGTGAACAGC 5460  
 Db 5401 CCGTGTGTGTGCGGCTCTTCAAGAGCTTCCCTGTCTCAATCTGAGAGTGAACAGC 5460  
 QY 5461 GCGAGGACGCGAGAGATCTGTGTGTTCAGAAATTTGAGATGTGTGTGATCT 5520  
 Db 5461 GCGAGGACGCGAGAGATCTGTGTGTTCAGAAATTTGAGATGTGTGTGATCT 5520  
 QY 5521 TAGGAAAGATGACCGGACAGAGATCTCAAGTGAAGTGTGCTTACCTTGGCTTGGC 5580  
 Db 5521 TAGGAAAGATGACCGGACAGAGATCTCAAGTGAAGTGTGCTTACCTTGGCTTGGC 5580  
 QY 5581 TACAGAGAACCTATCTGTGTGTGACCACTTCACTCACTGAGATTAATGAGATCAG 5640  
 Db 5581 TACAGAGAACCTATCTGTGTGTGACCACTTCACTCACTGAGATTAATGAGATCAG 5640  
 QY 5641 GCGAGCTCTGAGAGAGGAGCCCGGCGGAGCTTGAATCTGAGATCTGAGATCT 5700  
 Db 5641 GCGAGCTCTGAGAGAGGAGCCCGGCGGAGCTTGAATCTGAGATCTGAGATCT 5700  
 QY 5701 CTGGGCTCTGAGAGGAGGAGCCCGGCGGAGCTTGAATCTGAGATCTGAGATCT 5760  
 Db 5701 CTGGGCTCTGAGAGGAGGAGCCCGGCGGAGCTTGAATCTGAGATCTGAGATCT 5760  
 QY 5761 AAGGTATTTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5820  
 Db 5761 AAGGTATTTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5820  
 QY 5821 AGGTCATTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5880  
 Db 5821 AGGTCATTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5880  
 QY 5881 ACCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5940  
 Db 5881 ACCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5940  
 QY 5941 AGCAACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6000  
 Db 5941 AGCAACCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6000  
 QY 6001 CGCCCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6060  
 Db 6001 CGCCCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6060  
 QY 6061 CCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6120  
 Db 6061 CCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6120  
 QY 6121 CCGTGTCTTCAAGTGAACAGGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6180  
 Db 6121 CCGTGTCTTCAAGTGAACAGGATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6180  
 RESULT 4



AX671108  
LOCUS AX671108 8603 bp DNA linear PAT 27-MAR-2003  
DEFINITION Sequence 4 from Patent WO03004523.  
ACCESSION AX671108  
VERSION AX671108.1 GI:29329570  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 Zhu, Z.  
AUTHORS Regulation of human cItrn rho/rac-interacting kinase  
TITLE Patent: WO 03004523-A 4 16-JAN-2003;  
JOURNAL Bayer Aktienselschaft (DE)  
FEATURES  
Location/Qualifiers  
1..8603  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 2305 a 2206 c 2215 g 1877 t  
ORIGIN  
Query Match 99.9%; Score 6160.2; DB 6; Length 8603;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6162; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATGTGAAGTCAAAATATGAGCGCGGAAATCTTTGGATGCTGCTGTAACCAT 60  
Db 1 ATGTGAAGTCAAAATATGAGCGCGGAAATCTTTGGATGCTGCTGTAACCAT 60  
QY 61 GCCAGCGGGGCTCCAGGCTGATCTCTTCCAGGGGAAACACCCCTTATGACTCA 120  
Db 61 GCCAGCGGGGCTCCAGGCTGATCTCTTCCAGGGGAAACACCCCTTATGACTCA 120  
QY 121 CAGCAGATGTCCTCTTCCCGAGAGGATATAGATGCCCTTGTCTCTTGA 180  
Db 121 CAGCAGATGTCCTCTTCCCGAGAGGATATAGATGCCCTTGTCTCTTGA 180  
QY 181 GAATGACGTCAGCTGCTGATGATTAAGACGTCGACAACTTGTCCGAAAT 240  
Db 181 GAATGACGTCAGCTGCTGATGATTAAGACGTCGACAACTTGTCCGAAAT 240  
QY 241 TCCGACACATAGCTGATTAAGAGCTCCAGCTCCGCAAGGCTTCCAGTCA 300  
Db 241 TCCGACACATAGCTGATTAAGAGCTCCAGCTCCGCAAGGCTTCCAGTCA 300  
QY 301 AGCTTGTAGGTTGTGTCATCTTGTGTAAGTGAGGTGTAAGAGCAACCGG 360  
Db 301 AGCTTGTAGGTTGTGTCATCTTGTGTAAGTGAGGTGTAAGAGCAACCGG 360  
QY 361 GACATCATCTATGAAAGTGAAGAAAGAGGCTTATTTGGCCGAGAGGTTCA 420  
Db 361 GACATCATCTATGAAAGTGAAGAAAGAGGCTTATTTGGCCGAGAGGTTCA 420  
QY 421 TTTTGTAGGAAGCGGACATATTTATCTGAAGCAAGCCCGTGATCCCCAAT 480  
Db 421 TTTTGTAGGAAGCGGACATATTTATCTGAAGCAAGCCCGTGATCCCCAAT 480  
QY 481 CAGTATGCTTTGAGCAAAAATCACCTTTATCTGTCATGGAATATCAGCTGAG 540  
Db 481 CAGTATGCTTTGAGCAAAAATCACCTTTATCTGTCATGGAATATCAGCTGAG 540  
QY 541 GACTTGCTGCACTTTGATATGATGAGCAAGGATGATGAAAACCTGATACGTT 600  
Db 541 GACTTGCTGCACTTTGATATGATGAGCAAGGATGATGAAAACCTGATACGTT 600  
QY 601 TACCTAGCTAGCTATTTGGCTTCAAGCGTTATCTGATGAGGATACGTCAG 660  
Db 601 TACCTAGCTAGCTATTTGGCTTCAAGCGTTATCTGATGAGGATACGTCAG 660  
QY 661 GACATCAAGCTGAGAACTTCTGTTGACCGACAGACACATCAAGCGTGGATTT 720  
Db 661 GACATCAAGCTGAGAACTTCTGTTGACCGACAGACACATCAAGCGTGGATTT 720

Db 661 GACATCAAGCTGAGAACTTCTGTTGACCGACAGACACATCAAGCTGGATTT 720  
QY 721 GATTCGCGCGGAAATGATTTCAACAGATGTTGAATGCCAACTCCGATTTGGAC 780  
Db 721 GATTCGCGCGGAAATGATTTCAACAGATGTTGAATGCCAACTCCGATTTGGAC 780  
QY 781 CCAGATTCATGCTCTGAAGTGTGATGTAAGCGGGGATGAAAAAGGACCTAC 840  
Db 781 CCAGATTCATGCTCTGAAGTGTGATGTAAGCGGGGATGAAAAAGGACCTAC 840  
QY 841 GGCCTGACCTGTGACTGTGTGCTGAGTGGCGGTGATTTGCTATGAGATTTATGG 900  
Db 841 GGCCTGACCTGTGACTGTGTGCTGAGTGGCGGTGATTTGCTATGAGATTTATGG 900  
QY 901 TCCCGCTTCCGAGAGGAACTCTGCGCAACCTTCAATATATGATTTCCAGCG 960  
Db 901 TCCCGCTTCCGAGAGGAACTCTGCGCAACCTTCAATATATGATTTCCAGCG 960  
QY 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTCTTCTGATCTGATCAAG 1020  
Db 961 TTTTGAATTTCCAGATGACCCCAAGTGAAGTGAAGTCTTCTGATCTGATCAAG 1020  
QY 1021 TTTGTTGCGGCGCAGAAAGAGACTGAAGTTGAAGTCTTGTGCTGATCTTCTTC 1080  
Db 1021 TTTGTTGCGGCGCAGAAAGAGACTGAAGTTGAAGTCTTGTGCTGATCTTCTTC 1080  
QY 1081 TCTAAATTTGACTGGAACAATCTGTAACCTCTCCCTCCCTGTTCCACCTCAAG 1140  
Db 1081 TCTAAATTTGACTGGAACAATCTGTAACCTCTCTCCCTCTGTTCCACCTCAAG 1140  
QY 1141 TCTGACGATGACCTCCAAATTTGATGAACCAAGAAATTCGTGGTTTCACTCT 1200  
Db 1141 TCTGACGATGACCTCCAAATTTGATGAACCAAGAAATTCGTGGTTTCACTCT 1200  
QY 1201 CCGTGCAGCTGAGGCGCTCCAGGCTTCCGAGTGAAGAACTGCGGCTTGTGG 1260  
Db 1201 CCGTGCAGCTGAGGCGCTCCAGGCTTCCGAGTGAAGAACTGCGGCTTGTGG 1260  
QY 1261 TACAGCAAGCACTGGGATTTCTTGATGATCTGATGTTGTGTCGGATCTGAC 1320  
Db 1261 TACAGCAAGCACTGGGATTTCTTGATGATCTGATGTTGTGTCGGATCTGAC 1320  
QY 1321 CCGTGCAGCTGAGCTCCAGTGAAGAAAGAACTTCTCAACAAAGAGCTACAAAC 1380  
Db 1321 CCGTGCAGCTGAGCTCCAGTGAAGAAAGAACTTCTCAACAAAGAGCTACAAAC 1380  
QY 1381 TCTCAGGACCAAGTGTCAAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
Db 1381 TCTCAGGACCAAGTGTCAAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
QY 1441 GAGGTGAGGCTGTGCTTATGTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
Db 1441 GAGGTGAGGCTGTGCTTATGTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
QY 1501 TCCCTCTGAGAGCAGGACCTTGCTACCTCACTACAGAAATCAGTATTAAGGAG 1560  
Db 1501 TCCCTCTGAGAGCAGGACCTTGCTACCTCACTACAGAAATCAGTATTAAGGAG 1560  
QY 1561 TTTGAGCAGACCGATGAGGATGCTCCAGAGGATGCAAGACCTGCACTTCCAT 1620  
Db 1561 TTTGAGCAGACCGATGAGGATGCTCCAGAGGATGCAAGACCTGCACTTCCAT 1620  
QY 1621 GATATCAAGAGCAGAGCGGAGCTCCAAATTAAGAGCAGAGATCAAGAGCTTCA 1680  
Db 1621 GATATCAAGAGCAGAGCGGAGCTCCAAATTAAGAGCAGAGATCAAGAGCTTCA 1680  
QY 1681 GTGAGAAATGAGGTTGATGATCACTTGAAGAGATCTTGTCTCAGCAAGAGA 1740  
Db 1681 GTGAGAAATGAGGTTGATGATCACTTGAAGAGATCTTGTCTCAGCAAGAGA 1740  
QY 1741 CGGAGTGAATCTCTGAGATCTGAGCTGAGAGAGCTCGGCTTGTCTGAGAAATTC 1800  
Db 1741 CGGAGTGAATCTCTGAGATCTGAGCTGAGAGAGCTCGGCTTGTCTGAGAAATTC 1800



QY 1801 CGGAAAGCGACAGATGTGACGATTAACCTGTTGAAGCTAAGAGATCGAAGGAGCTGAA 1860  
 DB 1801 CGGAAAGCGACAGAAATGTGACGATTAACCTGTTGAAGCTAAGAGATCGAAGGAGCTGAA 1860  
 QY 1861 GTGGGGAATATGCGGAACTGGAGAAATGCAATGCTGAGCAGACCTCAAAATTCAGAGAG 1920  
 DB 1861 GTGGGGAATATGCGGAACTGGAGAAATGCAATGCTGAGCAGACCTCAAAATTCAGAGAG 1920  
 QY 1921 CTCGAAAGAGAACTGGAGAAAGCTGTAAAGCCAGACCGAGAGCCAGAGCTGTGAG 1980  
 DB 1921 CTCGAAAGAGAACTGGAGAAAGCTGTAAAGCCAGACCGAGAGCCAGAGCTGTGAG 1980  
 QY 1981 AATATCCCGCAGGCAAAAGAGCGAGCGAGGAGGAGCTGAGAAAGCTGCAAAACCGAGAG 2040  
 DB 1981 AATATCCCGCAGGCAAAAGAGCGAGCGAGGAGGAGCTGAGAAAGCTGCAAAACCGAGAG 2040  
 QY 2041 GATTCTTCTGAAGGCTCAGAAAGAGCTGTGAGAACTGAGAAAGCCGCTGCTCTG 2100  
 DB 2041 GATTCTTCTGAAGGCTCAGAAAGAGCTGTGAGAAAGCTGAGAAAGCCGCTGCTCTG 2100  
 QY 2101 GAGAACAGGTAAAGAGCTAAGAGACATGAGGCGTTAGAGAAACAGACTGAGAGTAGAC 2160  
 DB 2101 GAGAACAGGTAAAGAGCTAAGAGACATGAGGCGTTAGAGAAACAGACTGAGAGTAGAC 2160  
 QY 2161 ATCCAGACAAATCCCAACAGATCCAGAGATGAGCTGATTAATTTCTGAGCTCGAAGAG 2220  
 DB 2161 ATCCAGACAAATCCCAACAGATCCAGAGATGAGCTGATTAATTTCTGAGCTCGAAGAG 2220  
 QY 2221 AAACATCGGAGGCGCCAAAGTCTCAGCCAGCACTTGAAGTGCACCTGAAACAGAAAGAG 2280  
 DB 2221 AAACATCGGAGGCGCCAAAGTCTCAGCCAGCACTTGAAGTGCACCTGAAACAGAAAGAG 2280  
 QY 2281 CAGCACTATGAGGAGAAAGATTAAGTGTGAGCAATCAATTAAGAAAGACCTGCTGAC 2340  
 DB 2281 CAGCACTATGAGGAGAAAGATTAAGTGTGAGCAATCAATTAAGAAAGACCTGCTGAC 2340  
 QY 2341 AAGGAGACACTGAGAGAACTGATGAGAGACACAGAGAGAGGCCATAGAGAGGCAAA 2400  
 DB 2341 AAGGAGACACTGAGAGAACTGATGAGAGACACAGAGAGAGGCCATAGAGAGGCAAA 2400  
 QY 2401 ATTCTAGCGGAAACAGAGGCGATGATCAATGTCTATGAGATTCAGAAATTCAGATCCCTGAA 2460  
 DB 2401 ATTCTAGCGGAAACAGAGGCGATGATCAATGTCTATGAGATTCAGAAATTCAGATCCCTGAA 2460  
 QY 2461 CAGAGAGATTGTGAACTGTCTGAGAGCCAAATTAATTTGACAGAAATAGCAGTCTTTTACC 2520  
 DB 2461 CAGAGAGATTGTGAACTGTCTGAGAGCCAAATTAATTTGACAGAAATAGCAGTCTTTTACC 2520  
 QY 2521 CAAAGGACATGAGAGGCCCAAGAGAGATTTCTGAACTCAGGCAACAGAAATTTTAC 2580  
 DB 2521 CAAAGGACATGAGAGGCCCAAGAGAGATTTCTGAACTCAGGCAACAGAAATTTTAC 2580  
 QY 2581 CTGAGAGACAGAGCTGGAGAGTTGAGAGCCCAAGAACCAAACTGAGAGAGAGCTGGAG 2640  
 DB 2581 CTGAGAGACAGAGCTGGAGAGTTGAGAGCCCAAGAACCAAACTGAGAGAGAGCTGGAG 2640  
 QY 2641 AAGATAGCCCAAGACCAAGATCAAGTGAACAGATCGGCTGCTGAACTGAGAGCAAGATTG 2700  
 DB 2641 AAGATAGCCCAAGACCAAGATCAAGTGAACAGATCGGCTGCTGAACTGAGAGCAAGATTG 2700  
 QY 2701 CGGAGAGTCAATGAGACAGAGAGCAAGAACTGAGAGCTCAGAGGCGAGCTCAGAG 2760  
 DB 2701 CGGAGAGTCAATGAGACAGAGAGCAAGAACTGAGAGCTCAGAGGCGAGCTCAGAG 2760  
 QY 2761 CTACAGCTCTCCCTGAGAGAGCGGAGTCAAGATTGACAGCCCTGAGAGCTGAGAGG 2820  
 DB 2761 CTACAGCTCTCCCTGAGAGAGCGGAGTCAAGATTGACAGCCCTGAGAGCTGAGAGG 2820  
 QY 2821 GCGCTGAGAGAGCGGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 2880  
 DB 2821 GCGCTGAGAGAGCGGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 2880

QY 2881 GAGAGAGAGATCCAGGACCTCAGGACATAGAGATGAATCCAGGCAATTTGATGCT 2940  
 DB 2881 GAGAGAGAGATCCAGGACCTCAGGACATAGAGATGAATCCAGGCAATTTGATGCT 2940  
 QY 2941 CTTGCTAAGCTGTATCTGATATCAAGACCTGAGAGAGAGCTTAACAGCTGAGAG 3000  
 DB 2941 CTTGCTAAGCTGTATCTGATATCAAGACCTGAGAGAGAGCTTAACAGCTGAGAG 3000  
 QY 3001 GACAAAGCTGAACTCAACAAACAAACCTTCTACTTGTCCAAACATCTGATGAGCTTCT 3060  
 DB 3001 GACAAAGCTGAACTCAACAAACAAACCTTCTACTTGTCCAAACATCTGATGAGCTTCT 3060  
 QY 3061 GCGCCCAACAGAGATTTGACAACTGCGAGAGTGAAGTGAACATCTCCGCGGAGATC 3120  
 DB 3061 GCGCCCAACAGAGATTTGACAACTGCGAGAGTGAAGTGAACATCTCCGCGGAGATC 3120  
 QY 3121 ACGGAGAGAGATGACGCTTACAGCCAGAGCAACAGATGAGAGCTTGAAGACCAAG 3180  
 DB 3121 ACGGAGAGAGATGACGCTTACAGCCAGAGCAACAGATGAGAGCTTGAAGACCAAG 3180  
 QY 3181 TGCAACATGCTGAGAGAAAGCTCAATGAGTTTGAAGGCTTAAACGATGAGCTGCTGAA 3240  
 DB 3181 TGCAACATGCTGAGAGAAAGCTCAATGAGTTTGAAGGCTTAAACGATGAGCTGCTGAA 3240  
 QY 3241 AAAGAGCGGAGTGGAGAGGCTGAGAGAGGCTGAGAGTATGAGAAATCCAGTTTGA 3300  
 DB 3241 AAAGAGCGGAGTGGAGAGGCTGAGAGAGGCTGAGAGTATGAGAAATCCAGTTTGA 3300  
 QY 3301 TGTGAGGTTGAGAGCTGACAGAGATGCTGACACCGAGAAACAGAGAGGCGAGAGCC 3360  
 DB 3301 TGTGAGGTTGAGAGCTGACAGAGATGCTGACACCGAGAAACAGAGAGGCGAGAGCC 3360  
 QY 3361 GATGAGGAGATCAACCGAGTCTGCGAGAGTGGAGTGGAGCTGAGAGAGGCAAGAGCT 3420  
 DB 3361 GATGAGGAGATCAACCGAGTCTGCGAGAGTGGAGTGGAGCTGAGAGAGGCAAGAGCT 3420  
 QY 3421 GAGATTTCTGCTGACAGAGGCTCTCAAAAGACAGAAAGCTGAGAGGCGAGAGCTCTCT 3480  
 DB 3421 GAGATTTCTGCTGACAGAGGCTCTCAAAAGACAGAAAGCTGAGAGGCGAGAGCTCTCT 3480  
 QY 3481 GACAGAGTCAATGACCTGAGAGAAAGACAGCTATGCTTAATTAATGAGTGGCCCAAGCTTA 3540  
 DB 3481 GACAGAGTCAATGACCTGAGAGAAAGACAGCTATGCTTAATTAATGAGTGGCCCAAGCTTA 3540  
 QY 3541 CAGCAGAGAGTGGAGACTGAGAGAGAGCTCAAAACAGAGCTTCTGAGAGCAAGCCAAA 3600  
 DB 3541 CAGCAGAGAGTGGAGACTGAGAGAGAGCTCAAAACAGAGCTTCTGAGAGCAAGCCAAA 3600  
 QY 3601 TTACAGAGAGATGAGACCTGACAGAAATCAATTTTCCGTCTGATCAAGAGCTGCA 3660  
 DB 3601 TTACAGAGAGATGAGACCTGACAGAAATCAATTTTCCGTCTGATCAAGAGCTGCA 3660  
 QY 3661 GAAAGCTCTAGAGTGGAGTCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCA 3720  
 DB 3661 GAAAGCTCTAGAGTGGAGTCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCA 3720  
 QY 3721 GAAAAATTTGAGGTTCTGATTTCTCATGAAAGGTGAAAAATGAGAGCAATTTTCTCAA 3780  
 DB 3721 GAAAAATTTGAGGTTCTGATTTCTCATGAAAGGTGAAAAATGAGAGCAATTTTCTCAA 3780  
 QY 3781 CAAACCAAGCTCAATGATTTTCTGCAAGAGCAAAATGAGAGCAACCTGCTTAAGAAAG 3840  
 DB 3781 CAAACCAAGCTCAATGATTTTCTGCAAGAGCAAAATGAGAGCAACCTGCTTAAGAAAG 3840  
 QY 3841 GTTCTCTGAGATCAATGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGT 3900  
 DB 3841 GTTCTCTGAGATCAATGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGT 3900  
 QY 3901 GAGCTTAAAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 3960  
 DB 3901 GAGCTTAAAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 3960  
 QY 3961 GCGCAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 4020



RESULT 5  
 AX671112 6156 bp DNA linear PAT 27-MAR-2003  
 LOCUS  
 DEFINITION Sequence 8 from Patent WO03004523.  
 AX671112  
 ACCSSION  
 VERSION AX671112.1 GI:29329572  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE  
 AUTHORS Zhu Z.  
 TITLE Regulation of human citron rho/rac-interacting kinase  
 JOURNAL Patent: WO 03004523-A 8 16-JAN-2003;  
 BAYER Aktiengesellschaft (DE)  
 FEATURES  
 source 1..6156  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 1732 a 1548 c 1679 g 1197 t  
 ORIGIN  
 Query Match 99.6%; Score 6142.2; DB 6; Length 6156;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 6147; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 ATGTTGAAGTTCAATATGAGAGCGGGAATCCTTTGGATGCTGTGCTGTAACCAATT 60  
 DB 1 ATGTTGAAGTTCAATATGAGAGCGGGAATCCTTTGGATGCTGTGCTGTAACCAATT 60  
 QY 61 GCCAGCCGGGCGCTCCAGGCTGAATCTGTTCTTCCAGAGGAGAAACAACCTTTATGACTCA 120  
 DB 61 GCCAGCCGGGCGCTCCAGGCTGAATCTGTTCTTCCAGAGGAGAAACAACCTTTATGACTCA 120  
 QY 121 CAGCAGATGCTCTCTTCCGAGAGAGGATATGATGCCCTCTTGTCTTTGAA 180  
 DB 121 CAGCAGATGCTCTCTTCCGAGAGAGGATATGATGCCCTCTTGTCTTTGAA 180  
 QY 181 GATGAGTCAGCTGCTCTGATGAGATTAAGCACTGAGCACTTGTCCGGAAGTAT 240  
 DB 181 GATGAGTCAGCTGCTCTGATGAGATTAAGCACTGAGCACTTGTCCGGAAGTAT 240  
 QY 241 TCCGACACCATAGCTAGTTACAGAGCTCCAGCCTTCGCAAGAGACTTCGAGTCA 300  
 DB 241 TCCGACACCATAGCTAGTTACAGAGCTCCAGCCTTCGCAAGAGACTTCGAGTCA 300  
 QY 301 AGCTCTGTAGTGTGTGTCTTGTCTGTAAGTCAGAGTGTAAAGAGAAACCAACCGGG 360  
 DB 301 AGCTCTGTAGTGTGTGTCTTGTCTGTAAGTCAGAGTGTAAAGAGAAACCAACCGGG 360  
 QY 361 GACATCTATGCTATGAAGATGAGAGAGAGGCTTTATTTGGCCAGAGAGAGGTTTCA 420  
 DB 361 GACATCTATGCTATGAAGATGAGAGAGAGGCTTTATTTGGCCAGAGAGAGGTTTCA 420  
 QY 421 TTTTGGAGAGAGAGCGGAACATATTTATTCGAAGCAACGCCCTGTGATCCCAATTA 480  
 DB 421 TTTTGGAGAGAGAGCGGAACATATTTATTCGAAGCAACGCCCTGTGATCCCAATTA 480  
 QY 481 CAGTATGCTCTTTCAGAGCAAAAATCACTTTATCTGTGATGAGAAATATCAGGCTGAGGG 540  
 DB 481 CAGTATGCTCTTTCAGAGCAAAAATCACTTTATCTGTGATGAGAAATATCAGGCTGAGGG 540  
 QY 541 GACTTCTGTCTCACTTTTGAATGATATGAGACCAAGTTAGTAAACCTGATACAGTTT 600  
 DB 541 GACTTCTGTCTCACTTTTGAATGATATGAGACCAAGTTAGTAAACCTGATACAGTTT 600  
 QY 601 TACCTGAGTGAAGTATTTGGCTGTTCAAGGCTCATCTGATGGAGTAACGTGATGCA 660  
 DB 601 TACCTGAGTGAAGTATTTGGCTGTTCAAGGCTCATCTGATGGAGTAACGTGATGCA 660

QY 661 GACATCAAGCTTGAAGAACATTTCTGTGACCCGACAGAGACATCAAGCTGTGATTTT 720  
 DB 661 GACATCAAGCTTGAAGAACATTTCTGTGACCCGACAGAGACATCAAGCTGTGATTTT 720  
 QY 721 GGATCTGCGGGAAGATGAATTCAAACAAATGATGTAATCCAAACCTCCGATTTGGAGC 780  
 DB 721 GGATCTGCGGGAAGATGAATTCAAACAAATGATGTAATCCAAACCTCCGATTTGGAGC 780  
 QY 781 CCAATATTAATGAGCTCTGAGAGTGTGATGAGACGAGAGATGAGAAAGCACTAC 840  
 DB 781 CCAATATTAATGAGCTCTGAGAGTGTGATGAGACGAGAGATGAGAAAGCACTAC 840  
 QY 841 GGCCTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 DB 841 GGCCTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 QY 901 TCCCTCTTCCAGAGGAGAACCTTCGAGAGACCTTCATTAATTAATTTTCCAGCGG 960  
 DB 901 TCCCTCTTCCAGAGGAGAACCTTCGAGAGACCTTCATTAATTAATTTTCCAGCGG 960  
 QY 961 TTTTGAATTTTCCAGATGACCCCAAGATGAGAGTACTTTTGTATCTGATTCAGAGC 1020  
 DB 961 TTTTGAATTTTCCAGATGACCCCAAGATGAGAGTACTTTTGTATCTGATTCAGAGC 1020  
 QY 1021 TTTGTTGCGGCGAGAAAGAGAGAGTGAAGTGTGATGATGATGATGATGATGATGATGAT 1080  
 DB 1021 TTTGTTGCGGCGAGAAAGAGAGAGTGAAGTGTGATGATGATGATGATGATGATGATGAT 1080  
 QY 1081 TCTTAAATTTGATGAGAAACATTTCTGTAATCTTCTCCCTTCTGTTCCACCTTAAG 1140  
 DB 1081 TCTTAAATTTGATGAGAAACATTTCTGTAATCTTCTCCCTTCTGTTCCACCTTAAG 1140  
 QY 1141 TCTGAGATGACACTTCCATTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 DB 1141 TCTGAGATGACACTTCCATTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
 QY 1201 CCGTGCCAGCTGAGCCCTTCAAGCTTCTGAGGATGAGAACTCCGTTTGTGGGTTTTCG 1260  
 DB 1201 CCGTGCCAGCTGAGCCCTTCAAGCTTCTGAGGATGAGAACTCCGTTTGTGGGTTTTCG 1260  
 QY 1261 TAGAGCAAGGACCTGGGAGATTTCTGTGATGATGATGATGATGATGATGATGATGATGAT 1320  
 DB 1261 TAGAGCAAGGACCTGGGAGATTTCTGTGATGATGATGATGATGATGATGATGATGATGAT 1320  
 QY 1321 CCTGCAAGACTTACCTCCATGAGAAAGAACTTCTCATCAAAAGCAAGAGCTTACAGAGC 1380  
 DB 1321 CCTGCAAGACTTACCTCCATGAGAAAGAACTTCTCATCAAAAGCAAGAGCTTACAGAGC 1380  
 QY 1381 TCTTGAAGCAAGTGTCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 DB 1381 TCTTGAAGCAAGTGTCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 QY 1441 GAGGTGAGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
 DB 1441 GAGGTGAGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
 QY 1501 TCCCTCTTCCAGAGGAGAGCTTGTGATCACTATCAAGAAAGCAAGTAAAGCCAAAGT 1560  
 DB 1501 TCCCTCTTCCAGAGGAGAGCTTGTGATCACTATCAAGAAAGCAAGTAAAGCCAAAGT 1560  
 QY 1561 TTTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 DB 1561 TTTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 QY 1621 GATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
 DB 1621 GATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
 QY 1681 GTGGAAGAAATGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
 DB 1681 GTGGAAGAAATGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740

OY	1741	UGAGTGAATCTCTAAGAACTCTGAGTGAAGAAGTCTGGGCTTGCTGCTGAAGAATTCAAG	1800
Db	1741	CGAGTGAATCTCTAAGAACTCTGAGTGAAGAAGTCTGGGCTTGCTGCTGAAGAATTCAAG	1800
OY	1801	CGGAAGGCAAGAAATGTCAGACATAAATCTGTTGAAGGTAAAGATCAAGGAACCTGAA	1860
Db	1801	CGGAAGGCAAGAAATGTCAGCAATAAATCTGTTGAAGGTAAAGATCAAGGAACCTGAA	1860
OY	1861	GTGGAGAAATATGCGAAACTGAGAGATTCATGCTGAGCAAGCTCAAAATTCAGAG	1920
Db	1861	GTGGAGAAATATGCGAAAATCTGAGAGATTCATGCTGAGCAAGCTCAAAATTCAGAG	1920
OY	1921	CTCCAAGGAATCTGAGAGGCTGTAAATACCCAGCAAGAGGCCACCGAGCTGCTGAG	1980
Db	1921	CTCCAAGGAATCTGAGAGGCTGTAAATACCCAGCAAGAGGCCACCGAGCTGCTGAG	1980
OY	1981	AATATCCGCCAAGCAAGAGCGAGCCGAGAGGAGACTGGAAGACTTGAGAACTTGACAG	2040
Db	1981	AATATCCGCCAAGCAAGAGCGAGCCGAGAGGAGACTGGAAGACTTGAGAACTTGAGAG	2040
OY	2041	GATTCTTTGAAGGATCAAGAAAGAGTGTGTAAGGTGAGAAACCGCGCATTTCTG	2100
Db	2041	GATTCTTTGAAGGATCAAGAAAGAGTGTGTAAGGTGAGAAACCGCGCATTTCTG	2100
OY	2101	GAGAAACAAGTAAAGAGACTAGAGCACTGAGGCGTGAAGAAAAAGACTGAAGATGAC	2160
Db	2101	GAGAAACAAGTAAAGAGACTAGAGCACTGAGGCGTGAAGAAAAAGACTGAAGATGAC	2160
OY	2161	ATCCAGACAAATATCCCAACAGATCTCAGACAGATGCGTGAATAATTTGGAAGCTGAGAG	2220
Db	2161	ATCCAGACAAATATCCCAACAGATCTCAGAGAGATGCGTGAATAATTTGGAAGCTGAGAG	2220
OY	2221	AAACATCGGAGGCGCCAGTCTCAGCCGAGCACTGAGAGTGCACTTGAAACAGAAAG	2280
Db	2221	AAACATCGGAGGCGCCAGTCTCAGCCGAGCACTGAGAGTGCACTTGAAACAGAAAG	2280
OY	2281	CAGCACTATGAGGAAGAATTPAACTGTTGACAAATCAGATPAAAGAAAGCTGCTGAC	2340
Db	2281	CAGCACTATGAGGAAGAATTPAAAGTGTGAGCAATCAGATPAAAGAAAGCTGCTGAC	2340
OY	2341	AAGGAGACATCGGAGAACATATGTGAGAGACACGAGGAGAGGCCATGAGAAAGGCAAA	2400
Db	2341	AAGGAGACATCGGAGAACATATGTGAGAGACACGAGGAGAGGCCATGAGAAAGGCAAA	2400
OY	2401	ATTCTCAGCGAACAGAAAGCGCATGATCAATGCTATGAGATTCCAAGATCAGATCCCTGAA	2460
Db	2401	ATTCTCAGCGAACAGAAAGCGCATGATCAATGCTATGAGATTCCAAGATCAGATCCCTGAA	2460
OY	2461	CAGAGGATTTGGGAATCTGTCAGAGCCATAAATCTGAGAGCAATGCAAGCTTTTAAAC	2520
Db	2461	CAGAGGATTTGGGAATCTGTCAGAGCCATAAATCTGAGAGCAATGCAAGCTTTTAAAC	2520
OY	2521	CAAAGGAACATGAGAGGCCCAAGAAAGATGATTTCTGAACTCAGGCAACAGAAATTTTAC	2580
Db	2521	CAAAGGAACATGAGAGGCCCAAGAAAGATGATTTCTGAACTCAGGCAACAGAAATTTTAC	2580
OY	2581	CTGAGAGACACAGGCTGGGAAGTTGAGAGCCCAAGAACCGAAAACTGAGAGGACAGTGGAG	2640
Db	2581	CTGAGAGACACAGGCTGGGAAGTTGAGAGCCCAAGAACCGAAAACTGAGAGGACAGTGGAG	2640
OY	2641	AAGATCAGCCCAAGAACCAAGTGAACAAGTAATGCGCTGCGAACTTGAACAAGATTTG	2700
Db	2641	AAGATCAGCCCAAGAACCAAGTGAACAAGTAATGCGCTGCGAACTTGAACAAGATTTG	2700
OY	2701	CGGAGGTCAAGTCTTAGAGACAGAGGACAGAACTGAGGCTCAAGCGCCAGACTCACAG	2760
Db	2701	CGGAGGTCAAGTCTTAGAGACAGAGGAGAGAACTGAGGCTCAAGCGCCAGACTCACAG	2760
OY	2761	CTACAGCTCTCCCTGCAAGAGCGAGTCAACGTTGACAGGCCCTGCAAGGCTGCACGGGCG	2820
Db	2761	CTACAGCTCTCTCCCTGCAAGAGCGAGTCAACGTTGACAGGCCCTGCAAGGCTGCACGGGCG	2820
OY	2821	GCCCTGAGAGACGACTTTCGCCAGGCGAAGACAGAGCTGGAAGAGACCAACAGAAACT	2880

Db	2821	GGCTTGGAGACCCAGCTTCGCCAGGCGAAGACAGCTTGGAGAGACCAACAGAGACT	2880
Qy	2881	GAAGAGGAGATCCAGGACCTCACGCGCATATGAGATGAAATCCAGCCGCAAAATTTGATGCT	2940
Db	2881	GAAGGAGGAGATCCAGGACCTCACGCGCATATGAGATGAAATCCAGCCGCAAAATTTGATGCT	2940
Qy	2941	CTTGTTAAACAGCTGTACTGTATATACAGACCTGGAGAGACGCTAAACCGAGTACCGAG	3000
Db	2941	CTTGTTAAACAGCTGTACTGTATATACAGACCTGGAGAGACGCTAAACCGAGTACCGAG	3000
Qy	3001	GACAAAGCTGAACCTCAACCAACAAACCTTCACTTGTCCAAACCAACTCGATGAGGCTTCT	3060
Db	3001	GACAAAGCTGAACCTCAACCAACAAACCTTCACTTGTCCAAACCAACTCGATGAGGCTTCT	3060
Qy	3061	GGCCGCAACGACGAGATTGTACAACTGCGAAGTGAAGTGAACATCTCCGCTGGAGATC	3120
Db	3061	GGCCGCAACGACGAGATTGTACAACTGCGAAGTGAAGTGAACATCTCCGCTGGAGATC	3120
Qy	3121	ACGGAACGAGAGATGCGAGCTTACACAGCCAGAGCAAAACGATGAGGCTCTGAACACAGC	3180
Db	3121	ACGGAACGAGAGATGCGAGCTTACACAGCCAGAGCAAAACGATGAGGCTCTGAACACAGC	3180
Qy	3181	TGCACCATGCTGAGAGAACAGGTCATGGAATTTGAGGCCCTTAAACGATGAGCTCTGAA	3240
Db	3181	TGCACCATGCTGAGAGAACAGGTCATGGAATTTGAGGCCCTTAAACGATGAGCTCTGAA	3240
Qy	3241	AAAGAGCGGACGTGGAGGCTCTGGAAGAGGCTCTGAGTGAATGAAGAAATCCGATTGAG	3300
Db	3241	AAAGAGCGGACGTGGAGGCTCTGGAAGAGGCTCTGAGTGAATGAAGAAATCCGATTGAG	3300
Qy	3301	TGTGGGTTTCGAGAGCTGACAGAGATCTGGACACCGAGAAACAGACACGGGCGAGAGCC	3360
Db	3301	TGTGGGTTTCGAGAGCTGACAGAGATCTGGACACCGAGAAACAGACACGGGCGAGAGCC	3360
Qy	3361	GATCAGCGGATCAACCGAGTCTTCGACAGGTGTGAGCTGACAGTGAAGAGGACMAAGGCT	3420
Db	3361	GATCAGCGGATCAACCGAGTCTTCGACAGGTGTGAGCTGACAGTGAAGAGGACMAAGGCT	3420
Qy	3421	GAGATTTTCGCTTCGACGACGAGGCTCTGAAGAGAGAACTGAAAGGCCGAGAGCTCTCT	3480
Db	3421	GAGATTTTCGCTTCGACGACGAGGCTCTGAAGAGAGAACTGAAAGGCCGAGAGCTCTCT	3480
Qy	3481	GACAAAGCTCAATGACCTGGAAGAGAAAGCATGCTTATGCTTGAATGAATGCCGAGGCTTA	3540
Db	3481	GACAAAGCTCAATGACCTGGAAGAGAAAGCATGCTTATGCTTGAATGAATGCCGAGGCTTA	3540
Qy	3541	CAGCAGAGAGCTGAGAGACTGGAACGAGAGCTCAACAGAGGCTCTGGAAGGCAAGCCAAA	3600
Db	3541	CAGCAGAGAGCTGAGAGACTGGAACGAGAGCTCAACAGAGGCTCTGGAAGGCAAGCCAAA	3600
Qy	3601	TTACAGCGACGATGAGCTTGCAGAAAAATCACATTTTCGCTGACTCAAGACTGCCAA	3660
Db	3601	TTACAGCGACGATGAGCTTGCAGAAAAATCACATTTTCGCTGACTCAAGACTGCCAA	3660
Qy	3661	GAAAGCTCTAAGTCCGGGCTGATCTACCTGAAGACAGAAAGAGTGACTTGGAGTACACTG	3720
Db	3661	GAAAGCTCTAAGTCCGGGCTGATCTACCTGAAGACAGAAAGAGTGACTTGGAGTACACTG	3720
Qy	3721	GAAGGATTCAGGTTCTCTATTTCTCAAGAAAGGTGAAGAGGACACATTTCTCA	3780
Db	3721	GAAGGATTCAGGTTCTCTATTTCTCAAGAAAGGTGAAGAGGACACATTTCTCA	3780
Qy	3781	CAAAACCAAACTCATTTGATTTTCTGCAAGCCAAAATGAGCCAACTGCTAAAAAAGAAAAG	3840
Db	3781	CAAAACCAAACTCATTTGATTTTCTGCAAGCCAAAATGAGCCAACTGCTAAAAAAGAAAAG	3840
Qy	3841	GTTCCCTGACGTACAATGAGCTGAAGTGGCCCTGGAGAGAGGAGAAAGCTCGCTGAC	3900
Db	3841	GTTCCCTGACGTACAATGAGCTGAAGTGGCCCTGGAGAGAGGAGAAAGCTCGCTGAC	3900
Qy	3901	GAGCTAAGAGAGCCCTTCAGAGAACCCGATGAGCTTCGGTCCGCCCGGAGAGAACT	3960
Db	3901	GAGCTAAGAGAGCCCTTCAGAGAACCCGATGAGCTTCGGTCCGCCCGGAGAGAACT	3960



[illegible]

QY 6121 CCGCTGTCCAGGTGAACAAGTCTGGACCAAGTC 6155  
 Db 6121 CCGCTGTCCAGGTGAACAAGTGAAGGCAAGTC 6155

RESULT 6  
 AX166510 6159 bp DNA linear PAT 22-JUN-2001  
 LOCUS Sequence 1 from Patent WO0138503.  
 DEFINITION AX166510  
 ACCESSION AX166510  
 VERSION AX166510.1 GI:14546855  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
 1 Ploman, G.D., Whyte, D., Manning, G.S., Sudarsanam, S.S., Martinez, R.,  
 Planagan, P. and Clary, D.S.  
 Novel human protein kinases and protein kinase-like enzymes  
 Patent: WO 0138503-A 1 31-MAY-2001.

FEATURES  
 Location/Qualifiers  
 1..6159  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

BASE COUNT 1732 a 1549 c 1680 g 1198 t  
 ORIGIN

Query Match 99.4%; Score 6126; DB 6; Length 6159;  
 Best Local Similarity 99.8%; Freq. No. 0;  
 Matches 6145; Conservative 0; Mismatches 10; Indels 3; Gaps 1;

QY 1 ATGTGAAGTCAATATATGAGGCGCGAAATCCCTTGTGATGCTGCTGCTGAACCAT 60  
 Db 1 ATGTGAAGTCAATATGAGGCGCGAAATCCCTTGTGATGCTGCTGCTGAACCAT 60

QY 61 GCCAGCCGGGCTCTCCAGGCTGAATCTGTTCTCCAGGGGAAACACCTTTATGACTCA 120  
 Db 61 GCCAGCCGGGCTCTCCAGGCTGAATCTGTTCTCCAGGGGAAACACCTTTATGACTCA 120

QY 121 CACGAGTGTCTCTCTCCAGGAGGAGATATGATGCCCTCTTGTCTCTTTGA 180  
 Db 121 CACGAGTGTCTCTCTCCAGGAGGAGATATGATGCCCTCTTGTCTCTTTGA 180

QY 121 CACGAGTGTCTCTCTCCAGGAGGAGATATGATGCCCTCTTGTCTCTTTGA 180  
 Db 121 CACGAGTGTCTCTCTCCAGGAGGAGATATGATGCCCTCTTGTCTCTTTGA 180

QY 181 GAATGCAAGTCAAGCTCTGATGAAGATTAGCAAGTGAAGCACTTCCGGAAGTG 240  
 Db 181 GAATGCAAGTCAAGCTCTGATGAAGATTAGCAAGTGAAGCACTTCCGGAAGTG 240

QY 238 TATTCGACACCATAGCTAGTACAGAGCTCCAGCTTCGCAAGGACTTCGAGTC 297  
 Db 241 TATTCGACACCATAGCTAGTACAGAGCTCCAGCTTCGCAAGGACTTCGAGTC 300

QY 298 AGAAGCTTGTAGGTTGTGTCTCACTTGTCTGAAGTCAAGTGTGTAGAGAGAAAGCAAC 357  
 Db 301 AGAAGCTTGTAGGTTGTGTCTCACTTGTCTGAAGTCAAGTGTGTAGAGAGAAAGCAAC 360

QY 358 GGGGACATCTATGCTATGAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 417  
 Db 361 GGGGACATCTATGCTATGAAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420

QY 418 TCAATTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 477  
 Db 421 TCAATTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480

QY 478 TTACAGTATGCTTTTCAAGAGCAAAATCACTTTATCTGCTCATGGAATATAGCTGGA 537  
 Db 481 TTACAGTATGCTTTTCAAGAGCAAAATCACTTTATCTGCTCATGGAATATAGCTGGA 540

QY 538 GGGGACCTGCTGCTCACTTTGAATAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597  
 Db 541 GGGGACCTGCTGCTCACTTTGAATAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600

QY 598 TTTACCTAGCTAGCTGATTTTGGCTGTCAAGCTGATCTGATGGGATACGTGAT 657  
 Db 601 TTTACCTAGCTAGCTGATTTTGGCTGTCAAGCTGATCTGATGGGATACGTGAT 660

QY 658 CGAGACATCAAGCTTGAGAAATCTCTGTTGACCGACAGAGACATCAAGCTGTGAT 717  
 Db 661 CGAGACATCAAGCTTGAGAAATCTCTGTTGACCGACAGAGACATCAAGCTGTGAT 720

QY 718 TTTGATCTGCTCGGAGAAATGAATTCAAACAAGTGTGAATGCGCAATCCCGATTGG 777  
 Db 721 TTTGATCTGCTCGGAGAAATGAATTCAAACAAGTGTGAATGCGCAATCCCGATTGG 780

QY 778 ACCCCAGATTACATGAGCTGCTGAAGTGTGATGATGATGATGATGATGATGATGATGAT 837  
 Db 781 ACCCCAGATTACATGAGCTGCTGAAGTGTGATGATGATGATGATGATGATGATGATGAT 840

QY 838 TAGGCTCTGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897  
 Db 841 TAGGCTCTGACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900

QY 898 AGATCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 957  
 Db 901 AGATCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

QY 958 CGGTTTTTGAATTTCCAGATGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017  
 Db 961 CGGTTTTTGAATTTCCAGATGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020

QY 1018 AGCTTTGTTGCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077  
 Db 1021 AGCTTTGTTGCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080

QY 1078 TTCTCTAATTTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1137  
 Db 1081 TTCTCTAATTTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140

QY 1138 AAGTCTGACATGACCTTCAATTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197  
 Db 1141 AAGTCTGACATGACCTTCAATTTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200

QY 1198 TCTCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
 Db 1201 TCTCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

QY 1258 TCGTACAGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317  
 Db 1261 TCGTACAGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320

QY 1318 TCCCTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377  
 Db 1321 TCCCTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380

QY 1378 GACTCTGAGCAAGTGTCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437  
 Db 1381 GACTCTGAGCAAGTGTCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440

QY 1438 TCAAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
 Db 1441 TCAAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500

QY 1498 AGATCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557  
 Db 1501 AGATCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560

QY 1558 AGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
 Db 1561 AGTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620

QY 1618 CATGATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677  
 Db 1621 CATGATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680











QY 3781 CAAACCAACTCATTTTCTGCAAGCCAAATGACCACTTGCTAAAGAAAAG 3840  
DB 3778 CAAACCAACTCATTTTCTGCAAGCCAAATGACCACTTGCTAAAGAAAAG 3837  
QY 3841 GTTCCTCTGACATGACATGAGCTGAGCTGACCTGAGAGAGAAAGCTGCTGCA 3900  
DB 3838 GTTCCTCTGACATGACATGAGCTGAGCTGAGCTGAGAGAGAAAGCTGCTGCA 3897  
QY 3901 GAGCTAGAGAGAGCTTCAAGAAACCGGATGAGCTCCGCTCCGAGAGAGAGCT 3960  
DB 3898 GAGCTAGAGAGAGCTTCAAGAAACCGGATGAGCTCCGCTCCGAGAGAGAGCT 3957  
QY 3961 GCGCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020  
DB 3958 GCGCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4017  
QY 4021 ATGCGCATGTCGCGCATGTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080  
DB 4018 ATGCGCATGTCGCGCATGTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4077  
QY 4081 GCGCGCGCATGTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140  
DB 4078 GCGCGCGCATGTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4137  
QY 4141 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200  
DB 4138 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4197  
QY 4201 ACAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4260  
DB 4198 ACAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4257  
QY 4261 GAATGTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4320  
DB 4258 GAATGTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4317  
QY 4321 CCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4380  
DB 4318 CCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4377  
QY 4381 GGTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4440  
DB 4378 GGTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4437  
QY 4441 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4500  
DB 4438 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4497  
QY 4501 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560  
DB 4498 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4557  
QY 4561 GAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4620  
DB 4558 GAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4617  
QY 4621 GCAAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
DB 4618 GCAAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4677  
QY 4681 ACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4740  
DB 4678 ACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4737  
QY 4741 TGGGTCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4800  
DB 4738 TGGGTCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4797  
QY 4801 GCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4860  
DB 4798 GCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4857  
QY 4861 ATGAACCTGACGCTGCTTCACTGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920

DB 4858 ATGAACCTGACGCTGCTTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4917  
QY 4921 TACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980  
DB 4918 TACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4977  
QY 4981 CAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5040  
DB 4978 CAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 5037  
QY 5041 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5100  
DB 5038 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5097  
QY 5101 CAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5160  
DB 5098 CAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5157  
QY 5161 GCGCAAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5220  
DB 5158 GCGCAAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5217  
QY 5221 CGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5280  
DB 5218 CGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5277  
QY 5281 TGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5340  
DB 5278 TGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5337  
QY 5341 ATGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5400  
DB 5338 ATGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5397  
QY 5401 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5460  
DB 5398 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5457  
QY 5461 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5520  
DB 5458 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5517  
QY 5521 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5580  
DB 5518 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5577  
QY 5581 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5640  
DB 5578 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5637  
QY 5641 GCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5700  
DB 5638 GCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5697  
QY 5701 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5760  
DB 5698 CTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5757  
QY 5761 AAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5820  
DB 5758 AAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5817  
QY 5821 CCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5880  
DB 5818 CCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5877  
QY 5881 ACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5940  
DB 5878 ACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5937  
QY 5941 AGCAACCTGACGCTGCTTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6000



Db 5938 AGACACCCACCGCTACCGGAGGGGCGGACCGAGCTGGCAGAGGACAAGTCTCTGAC 5997  
Qy 6001 CGCCCCCTGAGCGGAGAGAGTCCCCCGGCGGATATCTAGCAGCGGAGAGCGGTCC 6060  
Db 5998 CGCCCCCTGAGCGGAGAGAGTCCCCCGGCGGATATCTAGCAGCGGAGAGCGGTCC 6057  
Qy 6061 CCCGCGAGGCTTTTGAAGACAGACGAGGCGCGCTGCTGCGGAGCGGAGAGAGC 6120  
Db 6058 CCGCGAGGCTGTTTGAAGACAGACGAGGCGCGCTGCTGCGGAGCGGAGAGAGC 6117  
Qy 6121 CCGCTGTCCAGGTGAACAAGTCTGGGACCAAGTCTTCAATATTA 6165  
Db 6118 CCGCTGTCCAGGTGAACAAGTCTGGGACCAAGTCTTCAATATTA 6162

RESULT 8  
AX503778  
LOCUS AX503778 6201 bp DNA linear PAT 27-SEP-2002  
DEFINITION Sequence 8 from Patent WO226826.  
ACCESSION AX503778  
VERSION AX503778.1 GI:23385964  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
AUTHORS 1  
Gelach, V.L., Macdougall, J.R., Smithson, G., Miller, I., Stone, D.,  
Gunter, E., Ellerman, K., Grosse, W.M., Alsbrock, J.P., Lepley, D.M.,  
Burgess, C.E., Padigar, M., Kekuda, R., Spyrek, K.A., Leach, M.D. and  
Shimkets, R.A.  
Proteins and nucleic acids encoding same  
Patent: WO 0226826-A 8 04-APR-2002;  
Curagen Corporation (US)

TITLE  
JOURNAL  
FEATURES  
source  
1. .6201  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 1736 a 1552 c 1707 g 1206 t  
ORIGIN

Query Match 99.0%; Score 6105.4; DB 6; Length 6201;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 6131; Conservative 0; Mismatches 21; Indels 3; Gaps 1;

Qy 1 ATGTTGAAGTCAAAATATGAGCGCGGAATCTTTGATGCTGCTGTAACCAT 60  
Db 1 ATGTTGAAGTCAAAATATGAGCGCGGAATCTTTGATGCTGCTGTAACCAT 60  
Qy 61 GCCAGCGGGCTCCAGGCTGATCTCTTCCAGAGGAAACACCTTTATGACTCA 120  
Db 61 GCCAGCGGGCTCCAGGCTGATCTCTTCCAGAGGAAACACCTTTATGACTCA 120  
Qy 121 CAGCAGATGTCCTCTTCCAGAGGAGATATAGATGCCCTCTTCTTCTTGA 180  
Db 121 CAGCAGATGTCCTCTTCCAGAGGAGATATAGATGCCCTCTTCTTCTTGA 180  
Qy 181 GAATGCAAGTACGCTGCTGATGAAGATTAGCAGTGAACAATTTGTCGGAAT 240  
Db 181 GAATGCAAGTACGCTGCTGATGAAGATTAGCAGTGAACAATTTGTCGGAAT 240  
Qy 241 TCCGACACCATAGCTAGTACAGAGCTCCAGCTTCCGCAAGAGACTTCAGAGTCA 300  
Db 241 TCCGACACCATAGCTAGTACAGAGCTCCAGCTTCCGCAAGAGACTTCAGAGTCA 300  
Qy 301 AGTCTTGAAGTGTGTGCTCACTTGTGAAGTGAAGTGTGAAGAGAAACCGG 360  
Db 301 AGTCTTGAAGTGTGTGCTCACTTGTGAAGTGAAGTGTGAAGAGAAACCGG 360  
Qy 361 GACATCATGCTATGAAGAGATGAAGAGAGGCTTATTTGGCCAGAGAGAGTTTCA 420  
Db 361 GACATCATGCTATGAAGAGATGAAGAGAGGCTTATTTGGCCAGAGAGAGTTTCA 420

Qy 421 TTTTGTGAGAGAGACCGGAACATATATCTGAAAGCACAGCCGTGATCCCCAATTA 480  
Db 421 TTTTGTGAGAGAGACCGGAACATATATCTGAAAGCACAGCCGTGATCCCCAATTA 480  
Qy 481 CAGTATGCTTTACAGAGCAAAAATCACCTTATCTGGTCATGGAATATCAGCTGGAGG 540  
Db 481 CAGTATGCTTTACAGAGCAAAAATCACCTTATCTGGTCATGGAATATCAGCTGGAGG 540  
Qy 541 GACTTGCTGCACTTTGAATATGATAGAGACCAATTTGATGAAATCCTGATCAAGTTT 600  
Db 541 GACTTGCTGCACTTTGAATATGATAGAGACCAATTTGATGAAATCCTGATCAAGTTT 600  
Qy 601 TACCTAGCTGAGCTGATTTTGGCTGTCAAGGCTTCACTGATGAGATAGCTGATGGA 660  
Db 601 TACCTAGCTGAGCTGATTTTGGCTGTCAAGGCTTCACTGATGAGATAGCTGATGGA 660  
Qy 661 GACATCAAGCTGAGAACATCTGCTGTTGACCGCACAGACACATCAAGCTGATGATTT 720  
Db 661 GACATCAAGCTGAGAACATCTGCTGTTGACCGCACAGACACATCAAGCTGATGATTT 720  
Qy 721 GATTCGCGCGGAAATATGATCAACAAGTGTGAATGCAAACTCCGATTTGGAGC 780  
Db 721 GATTCGCGCGGAAATATGATCAACAAGTGTGAATGCAAACTCCGATTTGGAGC 780  
Qy 781 GATTCGCGCGGAAATATGATCAACAAGTGTGAATGCAAACTCCGATTTGGAGC 777  
Db 781 GATTCGCGCGGAAATATGATCAACAAGTGTGAATGCAAACTCCGATTTGGAGC 777  
Qy 778 CCAAGTTACATAGCTCTGAAAGTGTGATGTAACCGGAGATGAAAGAGCACTTAC 840  
Db 778 CCAAGTTACATAGCTCTGAAAGTGTGATGTAACCGGAGATGAAAGAGCACTTAC 840  
Qy 841 GAGCTGAGCTGATGCTGTGATGATGAGGCTGATGATGATGATGATGATGATGATG 900  
Db 841 GAGCTGAGCTGATGCTGTGATGATGAGGCTGATGATGATGATGATGATGATGATG 900  
Qy 901 TCCGCTTCCAGAGGAGAACTCTGCGCAAGAACTTCAATTAATTAATTAATTAATTA 960  
Db 901 TCCGCTTCCAGAGGAGAACTCTGCGCAAGAACTTCAATTAATTAATTAATTAATTA 960  
Qy 961 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020  
Db 961 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020  
Qy 1021 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1080  
Db 1021 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1080  
Qy 1081 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1140  
Db 1081 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1140  
Qy 1141 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1200  
Db 1141 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1200  
Qy 1201 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1260  
Db 1201 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1260  
Qy 1261 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1320  
Db 1261 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1320  
Qy 1321 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1380  
Db 1321 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1380  
Qy 1381 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1440  
Db 1381 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1440  
Qy 1441 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1500  
Db 1441 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1500  
Qy 1501 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1560  
Db 1501 TTTTGAATTTCCAGATGAGCCCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1560



QY	1501	TCCTCCCTGGAGAGCAAGCTTGCTCACTCAATCAAGAAATCAGATGAGCTTAAAGGAAGT	1560
Db	1498	TCCTCCCTGGAGAGCAAGCTTGCTCACTCAATCAAGAAATCAGATGAGCTTAAAGGAAGT	1557
QY	1561	TTGGAGCAGACCGAGTGAAGGTGTCCAGAGAGATGACAAAGCATCTGACCTTCCAT	1620
Db	1558	TTGGAGCAGACCGAGTGAAGGTGTCCAGAGAGATGACAAAGCATCTGACCTTCCAT	1617
QY	1621	GATATCAGAGAGCAGAGCCGGAGGTCCAGAAATCAAAGAGCAGAGATACAGGCTCAA	1680
Db	1618	GATATCAGAGAGCAGAGCCGGAGGTCCAGAAATCAAAGAGCAGAGATACAGGCTCAA	1677
QY	1681	GTGGAGAAATGAGGTTGATGATGAATCAATTGGAAGAGATCTTGTCTCAGACAGAGA	1740
Db	1678	GTGGAGAAATGAGGTTGATGATGAATCAATTGGAAGAGATCTTGTCTCAGACAGAGA	1737
QY	1741	CGAGATGATCTCTACGAATCTGAGTGAAGAGTCTCGGCTTGTGCTGAAGATTCAG	1800
Db	1738	CGAGATGATCTCTACGAATCTGAGTGAAGAGTCTCGGCTTGTGCTGAAGATTCAG	1797
QY	1801	CGGAAGCGAGACAGATGTCAAGCATMACTGTTGAAGGCTTAAGGATCAAGGAGAGCTGAA	1860
Db	1798	CGGAAGCGAGACAGATGTCAAGCATMACTGTTGAAGGCTTAAGGATCAAGGAGAGCTGAA	1857
QY	1861	GTGGAGAAATATCGGAAATCTGAGAGAGTCAATGCTGAGCAGCAGCTCAAAATTCAGAG	1920
Db	1858	GTGGAGAAATATCGGAAATCTGAGAGAGTCAATGCTGAGCAGCAGCTCAAAATTCAGAG	1917
QY	1921	CTCCAGAGAACTCTGAGAGAGGCTGTAAAGCCAGACAGAGAGCCAGACCTGCTGAG	1980
Db	1918	CTCCAGAGAACTCTGAGAGAGGCTGTAAAGCCAGACAGAGAGCCAGACCTGCTGAG	1977
QY	1961	AATATCGCCACAGGCAAGAGCAGACCCAGAGAGGAGCTTGAGAACTGACAGACCGAG	2040
Db	1978	AATATCGCCACAGGCAAGAGCAGACCCAGAGAGGAGCTTGAGAACTGACAGACCGAG	2037
QY	2041	GATCTCTGTAGAGCATGAGAAAGAGCTGTGGAAGCTGAGGAAAGCCGCCATCTCTG	2100
Db	2038	GATCTCTGTAGAGCATGAGAAAGAGCTGTGGAAGCTGAGGAAAGCCGCCATCTCTG	2097
QY	2101	GAGAAACAAGCTTAAAGAGACTAGAGACATGAGAGCTTAGAGAAAACAGACTAAGAGATAC	2160
Db	2098	GAGAAACAAGCTTAAAGAGACTAGAGACATGAGAGAGCTTAGAGAAAACAGACTAAGAGATAC	2157
QY	2161	ATCCAGACAAATCCCAACAGATCCAGAGATGCTGTATAATTTGAGACTCGAAGAG	2220
Db	2158	ATCCAGACAAATCCCAACAGATCCAGAGATGCTGTATAATTTGAGACTCGAAGAG	2217
QY	2221	AAACATCGGAGAGCCCAAGTCTCAGCCAGACCTAGAAAGTGCACCTGAACAGAAAG	2280
Db	2218	AAACATCGGAGAGCCCAAGTCTCAGCCAGACCTAGAAAGTGCACCTGAACAGAAAG	2277
QY	2281	CAGCATATGAGGAAGATTTAAAGTGTGGAACAATCAGATTAAGAAAGACTGCTGAC	2340
Db	2278	CAGCATATGAGGAAGATTTAAAGTATTTGACAAATCAGATTAAGAAAGACTGCTGAC	2337
QY	2341	AAGGAGACATCGAGAAACATGATGAGAGACACGAGAGAGAGGCCCATGAGAGGGCAA	2400
Db	2338	AAGGAGACATCGAGAGACATGATGAGAGACACGAGAGAGAGGCCCATGAGAGGGCAA	2397
QY	2401	ATTCTCAGCGAACAGAAAGCGCATGATCAATGCTATGGAATTCAGAGTCAATCCCTGGA	2460
Db	2398	ATTCTCAGCGAACAGAAAGCGCATGATCAATGCTATGGAATTCAGAGTCAATCCCTGGA	2457
QY	2461	CAGAGGATGTGGAACTGTCTGAAGCCAACTAACTTGAGAGAAATAGAGACTTTTAAAC	2520
Db	2458	CAGAGGATGTGGAACTGTCTGAAGCCAACTAACTTGAGAGAAATAGAGACTTTTAAAC	2517
QY	2521	CAAGAGAACATGAAGGCCCAAGAGAGATGATTTCTGAATCAAGGCAACAGAAATTTTAC	2580
Db	2518	CAAGAGAACATGAAGGCCCAAGAGAGATGATTTCTGAATCAAGGCAACAGAAATTTTAC	2577
QY	2581	CTGAGAGACAGAGGCTGGGAAGTTGAGAGCCAGAACCGAAATCTGAGAGAGCACTGGAG	2640

D	2578	CTGGAGACACAGCTGGGAAAGTTGGAGGCCAAGCCGAAACTGGAGGAGCAGCTGGAG	2637
Q	2641	AAGATCAGCCACCAAGACCAACAGTACAGAAATCGCTGGAACTGGAGACAGATTG	2700
D	2638	AAGATCAGCCACCAAGACCAACAGTACAGAAATCGCTGGAACTGGAGACAGATTG	2697
Q	2701	CGGAGGCTCAGCTCAGAGACCAAGGAGACCAAACTGGAGCTCAAGGCGCAGCTCACAG	2760
D	2698	CGGAGGCTCAGCTCAGAGACCAAGGAGACCAAACTGGAGCTCAAGGCGCAGCTCACAG	2757
Q	2761	CTACAGCTCTCCCTGCAGAGCGCGAGTCAcAGTTGACAGCCCTTCAGGCTGCACGAGCG	2820
D	2758	CTACAGCTCTCCCTGCAGAGCGCGAGTCAcAGTTGACAGCCCTTCAGGCTGCACGAGCG	2817
Q	2821	GCCTTGGAGAGCCAGCTTCCGCGAGGCGAAGACAGAGCTGGAAAGATCCACAGAAAGCT	2880
D	2818	GCCTTGGAGAGCCAGCTTCCGCGAGGCGAAGACAGAGCTGGAAAGAGACACAGAGAAAGCT	2877
Q	2881	GAAGAGAGATCCAGGCACTCACGGCACTATAGAGATAAAATCCAGCGCAATTTGATGCT	2940
D	2878	GAAGAGAGATCCAGGCACTCACGGCACTATAGAGATAAAATCCAGCGCAATTTGATGCT	2937
Q	2941	CTTCGTAAcAGCTGTACTGTATCAcAGACTTGAGAGGAGCAGCTTAAcCAGCTGACCGAG	3000
D	2938	CTTCGTAAcAGCTGTACTGTGTATCAcAGACTTGAGAGGAGCAGCTTAAcCAGCTGACCGAG	2997
Q	3001	GACAAcCCTGAATCAACCAACCAAAATTTCTATCTGTCCAACCAACTCGATGAGGCTCT	3060
D	2998	GACAAcCCTGAATCAACCAACCAAAATTTCTATCTGTCCAACCAACTCGATGAGGCTCT	3057
Q	3061	GGCGCCAACGACGAGATTGTAAcCTGCGAAGTGAATGAGCACTCCGCGCGGAGATC	3120
D	3058	GGCGCCAACGACGAGATTGTAAcCTGCGAAGTGAATGAGCACTCCGCGCGGAGATC	3117
Q	3121	ACGGAACGAGAGATGCGAGCTTACcAGCCACAAGCAAAcCATGAGAGCTCTGAGACCAcAG	3180
D	3118	ACGGAACGAGAGATGCGAGCTTACcAGCCACAAGCAAAcCATGAGAGCTCTGAGACCAcAG	3177
Q	3181	TGcACcATGCTGAGAGAAcAGGTCACTGGATTTGGAGGCGCTTAAcCATGAGCTCTGAA	3240
D	3178	TGcACcATGCTGAGAGAAcAGGTCACTGGATTTGGAGGCGCTTAAcCATGAGCTCTGAA	3237
Q	3241	AAAGAGCGGcAGTGGGAGGcCTTGAGAGGcCTCTGGGTATGAGAAATCCCACTTTGAG	3300
D	3238	AAAGAGCGGcAGTGGGAGGcCTTGAGAGGcCTCTGGGTATGAGAAATCCCACTTTGAG	3297
Q	3301	TGTCGGGTTGGAGAGCTGCGAGAGGATGCTGTGCAcCCGAGAAcCAGACGAGGAGAGCC	3360
D	3298	TGTCGGGTTGGAGAGCTGCGAGAGGATGCTGTGCAcCCGAGAAcCAGACGAGGAGAGCC	3357
Q	3361	GATCAGCGGATCAcCCAGTCTTCGcCAGGTGTGAGcCTGGcAGTGAAGAGCAcCAAGCT	3420
D	3358	GATCAGCGGATCAcCCAGTCTTCGcCAGGTGTGTGAGcCTGCAGTGAAGAGCAcCAAGCT	3417
Q	3421	GAGATTCTCGCTCGcAGcAGGCTTCCAAAGAGAGAGAGcGTGAAGGCGCGAGcCTCT	3480
D	3418	GAGATTCTCGCTCGcAGcAGGCTTCCAAAGAGAGAGAGcGTGAAGGCGCGAGcCTCT	3477
Q	3481	GACAAcGCTCAATGACCTGGAGAAAGcCATGTATGTTAAATGATGCGCCGAGCTT	3540
D	3478	GACAAcGCTCAATGACCTGGAGAAAGcCATGTATGTTAAATGATGCGCCGAGCTT	3537
Q	3541	CAGAGAAAGCTGAGAGCTGAAcCGAGAGCTTAAcCAGAGcCTTCTGGAAAGcCAAGCTCAA	3600
D	3538	CAGAGAAAGCTGAGAGCTGAAcCGAGAGCTTAAcCAGAGcCTTCTGGAAAGcCAAGCTCAA	3597
Q	3601	TTACAGACGACGATGAGcCTGCGAGAAAAATCAcATTTTCCGTCTGACTCAAGAGCTGCA	3660
D	3598	TTACAGACGACGATGAGcCTGCGAGAAAAATCAcATTTTCCGTCTGACTCAAGAGCTGCA	3657
Q	3661	GAAcCTTAAGTGGGCTGATCTTACTGAAAGcAGAAAGAGTGACTTGAAGTACAGCTG	3720

Dp	3658	GAAAGCTCAGATCCGGGCTGATGTTCTACTGAAAGACGAAAGAAAGTGAATCTTGAGATATGAGCTG	3712
Qy	3721	GAAAACATTCAAGTCTCTGATTTCTCATGAAAAGGTGAAAATGGAAGGACATATTTCTCAA	3780
Dp	3718	GAAAACATTCAAGTCTCTATTTCTCATGAAAAGGTGAAAATGGAAGGACATATTTCTCAA	3777
Qy	3781	CAAAACCAATCATATGATTTTCTGGAAGCGAAAATGAGCCAACTGCTTAAAAAGAAAAG	3840
Dp	3778	CAAAACCAATCATATGATTTTCTGGAAGCGAAAATGAGCCAACTGCTTAAAAAGAAAAG	3837
Qy	3841	GTTCCCTGCAAGTACATAGAGTGAAGCTGAGCCCTGAGAAAGAGAAAAGCTCGCTGTGCA	3900
Dp	3838	GTGCTCTGCAAGTACATAGAGTGAAGCTGAGCCCTGAGAAAGAGAAAAGCTCGCTGTGCA	3899
Qy	3901	GAGCTAGAGGAAGCCCTTCAGAAAGCCGCACTCGAGCTTCGGTCTGCCCGGAGAGAGCT	3960
Dp	3898	GAGCTAGAGGAAGCCCTTCAGAAAGCCGCACTCGAGCTTCGGTCTGCCCGGAGAGAGCT	3957
Qy	3961	GCCCCGCCCAAGCAACGAGGACCAACCCACATCCAGCCGACGACCGAGGAGAGAG	4020
Dp	3958	GCCCCGCCCAAGCAACGAGGACCAACCCACATCCAGCCGACGACCGAGGAGAGAG	4017
Qy	4021	ATGCGCATGTGCGCATGTGCGGATGCGCAAGACACAGCCCAATGCCATGAGCTGCTG	4080
Dp	4018	ATGCGCATGTGCGCATGTGCGGATGCGCAAGACACAGCCCAATGCCATGAGCTGCTG	4077
Qy	4081	GCCCCGCATCGACGCGGAGAAAGAGTCTTCAACTCGAGAGGAATTTAGTCGGGTCTT	4140
Dp	4078	GCCCCGCATCGACGCGGAGAAAGAGTCTTCAACTCGAGAGGAATTTAGTCGGGTCTT	4137
Qy	4141	AAGGAAGCATGACCAACATATTTCTCACCGATTCAACTGATGAGCTGAAATGAGAGCC	4200
Dp	4138	AAGGAAGCATGACCAACATATTTCTCACCGATTCAACTGATGAGCTGAAATGAGAGCC	4197
Qy	4201	ACAAAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4260
Dp	4198	ACAAAGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4257
Qy	4261	GAATGTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4320
Dp	4258	GAATGTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	4317
Qy	4321	CCTGCTGAATTTGCCACACACTTCCACGAGGCTTTCTGCGGTGACAAATGAACTCCCA	4380
Dp	4318	CCTGCTGAATTTGCCACACACTTCCACGAGGCTTTCTGCGGTGACAAATGAACTCCCA	4377
Qy	4381	GATCTCCAGACCAAGGAGCCCGACGACAGCTTGACCTTGAAAGGTGTGTAAGTGTGCC	4440
Dp	4378	GATCTCCAGACCAAGGAGCCCGACGACAGCTTGACCTTGAAAGGTGTGTAAGTGTGCC	4437
Qy	4441	AGGAATTAACAACGAGACAGCAAGCTGGAGACAGGAATACATTTGCTTGAAGATCA	4500
Dp	4438	AGGAATTAACAACGAGACAGCAAGCTGGAGACAGGAATACATTTGCTTGAAGATCA	4497
Qy	4501	AAAGTCCCATTTTGAATGAAGAGCCGAGAAAGTGTGACAGAGGCGGTGTGAAGATTT	4560
Dp	4498	AAAGTCCCATTTTGAATGAAGAGCCGAGAAAGTGTGACAGAGGCGGTGTGAAGATTT	4557
Qy	4561	GAGCTGTGCTTCCCGACGGGGAATGATATTTATGATGTGTGCGCTGTGTGCTTCGAACTC	4620
Dp	4558	GAGCTGTGCTTCCCGACGGGGAATGATATTTATGATGTGTGCGCTGTGTGCTTCGAACTC	4617
Qy	4621	GCAAAATACGCGAAGACGATGTCCCATATCATCTGAAGATGAAATCTCACCCGACACCC	4680
Dp	4618	GCAAAATACGCGAAGACGATGTCCCATATCATCTGAAGATGAAATCTCACCCGACACCC	4677
Qy	4681	ACCTGTGAGCCCGGAGAGACCTCTACTTGTGTAGCTCCACACTTCCCTGACAAACAGGCG	4740
Dp	4678	ACCTGTGAGCCCGGAGAGACCTCTACTTGTGTAGCTCCACACTTCCCTGACAAACAGGCG	4737
Qy	4741	TGGGTCAACCGCTTGAATCAATGTGTGCAAGTGTGGAAGATTTTGAAGAAAAGACGAA	4800
Dp	4738	TGGGTCAACCGCTTGAATCAATGTGTGCAAGTGTGGAAGATTTTGAAGAAAAGACGAA	4797

QY	4801	CGTAGTGTAAACGCTTGGAAATCCCTCGTGAACCTGGAAGGTGAAGACCGCTTAGAC	4860
Db	4798	GCTGATGTAAACGCTTGGAACTCCCTGTGAACCTGGAAGGTGAAGACCGCTTAGAC	4857
QY	4861	ATGAAGTCAAGCTGCCCTTCAGTGAACAGGTGTGTGGTGGACAACGAGGAAGGCTC	4920
Db	4858	ATGAAGTCAAGCTGCCCTTCAGTGAACAGGTGTGTGGTGGACAACGAGGAAGGCTC	4917
QY	4921	TACGCCCTGAAAGTCTTGAATAACTCCCTAACCCATGTCCAGGAATTGGAGCATCTTC	4980
Db	4918	TACGCCCTGAAAGTCTTGAATAACTCCCTAACCCATGTCCAGGAATTGGAGCATCTTC	4977
QY	4981	CAAAATTTATATATCAAGACCTTGAGAGTACTCATGATAGCAGAGGAAGACGGGACA	5040
Db	4978	CAAAATTTATATATCAAGACCTTGAGAGTACTCATGATAGCAGGTGAAGACGGGACA	5037
QY	5041	CTGTGCTCTTGGAGAGTGAATAAAGTGAACAGTCCCTGGCCCACTGGCTTGGC	5100
Db	5038	CTGTGCTCTTGGAGAGTGAATAAAGTGAACAGTCCCTGGCCCACTGGCTTGGC	5097
QY	5101	CAGCCCGACATCTCAACCCACATTTTGAAGCTGTCAAGGACTGCACTTGTGGGGACA	5160
Db	5098	CAGCCCGACATCTCAACCCACATTTTGAAGCTGTCAAGGACTGCACTTGTGGGGACA	5157
QY	5161	GGCAAGATTGGAAGAGGGCTCTGATCTGTGAGACCATGGCCAGCAAGTCTCATCTC	5220
Db	5158	GGCAAGATTGGAAGAGGGCTCTGATCTGTGAGACCATGGCCAGCAAGTCTCATCTC	5217
QY	5221	CGTACAAACGAAAACTTAGCAAAATCTGCAATCCGAGAAAGATAGAGACCTCAGAGCC	5280
Db	5218	CGCTACAAACGAAAACTTAGCAAAATCTGCAATCCGAGAAAGATAGAGACCTCAGAGCC	5277
QY	5281	TGCAAGCTATACCACTTCACCAATTACAGTATCCTCATTTGGAACCAATAATCTACGA	5340
Db	5278	TGCAAGCTATACCACTTCACCAATTACAGTATCCTCATTTGGAACCAATAATCTACGA	5337
QY	5341	ATCGACATGAAGAGTACACGCTTGAGGAATTCCTGTGATAAGATGACATTCCTTGGCA	5400
Db	5338	ATCGACATGAAGAGTACACGCTTGAGGAATTCCTGTGATAAGATGACATTCCTTGGCA	5397
QY	5401	CTGTGTGTGTGGCGGCTCTTCCAAAGAGTCCCTGTCAATCGTGGAGGTGAACAGC	5460
Db	5398	CTGTGTGTGTGGCGGCTCTTCCAAAGAGTCCCTGTGTCAATCGTGGAGGTGAACAGC	5457
QY	5461	GCAAGGGCAGCAGAGAGTACTTGTGTTCACGAATTTGGAGTTCGTGTGATTC	5520
Db	5458	GCAAGGGCAGCAGAGAGTACTTGTGTTCACGAATTTGGAGTTCGTGTGATTC	5517
QY	5521	TACGGAGAAGTACCCGACAGAGATCTGAAGTGAAGTGGCTTACTCTTGGGCTTTGCC	5580
Db	5518	TACGGAGAAGTACCCGACAGAGATCTGAAGTGAAGTGGCTTACTCTTGGGCTTTGCC	5577
QY	5581	TACAGAGAACCCTATTCTGTTTGTGACCCACTTCAACTCACTCGAAGTAATTGATCCAG	5640
Db	5578	TACAGAGAACCCTATTCTGTTTGTGACCCACTTCAACTCACTCGAAGTAATTGATCCAG	5637
QY	5641	GCAAGCTCTCAGAGGGAACCCCTGGCCCGAGCGTACTGGAACATCCGAAACCCGCGTAC	5700
Db	5638	GCAAGCTCTCAGAGGGAACCCCTGGCCCGAGCGTACTGGAACATCCGAAACCCGCGTAC	5697
QY	5701	CTGGAGCCCTGECATTTCTCAGAGAGCATTTACTTGGCGTCCATATCCAGGATTAATTA	5760
Db	5698	CTGGAGCCCTGECATTTCTCAGAGAGCATTTACTTGGCGTCCATATCCAGGATTAATTA	5757
QY	5761	AGGGTCAATTGCTGSCAAGGGAACCTCTGTAAGAGTCCGCACTGAACCAACCGGGGC	5820
Db	5758	AGGGTCAATTGCTGSCAAGGGAACCTCTGTAAGAGTCCGCACTGAACCAACCGGGGC	5817
QY	5821	CCGTTCACCTTCGCGACGAGCCCAACAGCAGAGGCCCAACCACTGACAGACATTC	5880
Db	5818	CCGTTCACCTTCGCGACGAGCCCAACAGCAGAGGCCCAACCACTGACAGACATTC	5877

QY 5881 ACCAAGCGGCTGCTCCAGCCCGCCCGCCGAGCCCGCCAGCCCGCCGAGAGCCA 5940  
 DB 5878 ACCAAGCGGCTGCTCCAGCCCGCCCGCCGAGCCCGCCAGCCCGCCGAGAGCCA 5937  
 QY 5941 AGACACCGCCCGCTACCGCGAGGGGCGGAGCCGAGCTCGGAGGAGCAAGCTCTGCGC 6000  
 DB 5938 AGACACCGCCCGCTACCGCGAGGGGCGGAGCCGAGCTCGGAGGAGCAAGCTCTGCGC 5997  
 QY 6001 CGCCCGCTGAGCGAGAGAGTCCCGCGCGGATCTCAGCAGCGCGAGAGAGCGGCTCC 6060  
 DB 5998 CGCCCGCTGAGCGAGAGAGTCCCGCGCGGATCTCAGCAGCGCGAGAGAGCGGCTCC 6057  
 QY 6061 CGCGAGGCTGTTTGAAGACAGCAGCGAGCGCGCGCTCGCGGAGCGCTGAGAGCC 6120  
 DB 6058 CGCGAGGCTGTTTGAAGACAGCAGCGAGCGCGCGCTCGCGGAGCGCTGAGAGCC 6117  
 QY 6121 CGCGCTGCTCCAGGTGACAGAGCTCTGCGAGCAGCTC 6155  
 DB 6118 CGCGCTGCTCCAGGTGACAGAGCTGAGCGAGCACTTC 6152

RESULT 9  
 AY257469 8576 bp mRNA linear PRI 23-APR-2003  
 LOCUS Homo sapiens rho/rac-interacting citron kinase (CIT) mRNA, complete cds.  
 DEFINITION  
 ACCESSION AY257469  
 VERSION AY257469.1 GI:30088969  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Huang, C.Q., Wu, S.L., Shan, Y.X., Liu, S. and Xiao, P.J.  
 TITLE Direct Submision  
 JOURNAL Submitted (18-MAR-2003) Department of Biochemistry and Molecular Biology, Basic Medicine, Suzhou University, Renmimg Road 48, Suzhou, Jiangsu 215007, China

FEATURES  
 source 1..8576  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="12"  
 /map="12q24.23"  
 1..8576  
 /gene="CIT"  
 /note="synonym: CRK"

CDS

/gene="CIT"  
 /note="serine/threonine protein kinases"  
 /product="rho/rac-interacting citron kinase"  
 /protein\_id="AA013528.1"  
 /db\_xref="GI:30088970"  
 /translation="MLKRYGARNPLDGAAPFASRLNLFQKPRPMTQOMS  
 PLSRGLDALTVEFGSPALMKIKHVSFVKSTIALEQLQISADQFVRL  
 VSGRFAEVQVREKATGDIYAMKVMKALLADQVQFPEERINLSRSPMPOL  
 QYAFQKNTLYVMYQPGDLSLNRVEDQDNLIOFYLAELILAVSHVGMGY  
 HDIKENILVDRTHIKLVDFGSAKMSNMVNAKPIGTPVAVBEVTVMGDG  
 KGTGLGCDMWSGVAVIYGRSPFAGTSARTFNNIMNORFLKPPDKVSDP  
 LDILOSLGOKERLKEBGLCOHPFSKIDMNNINSPPPPTPKSDDDSNPRE  
 KNSWSSPCQJSPSGFSELPVQFYSKXSLGILGSESVSGLDSPATSSMEKK  
 LTKSKLDSQKCHRMQENTRLRRSEVAVLSQKVEVLSKSTQKSLLEDLA  
 TYITKESLDSQKCHRMQENTRLRRSEVAVLSQKVEVLSKSTQKSLLEDLA  
 MNQLEEDVSAARRSDIYELRESRLAEFRKATCCQKILKADQKPEVGEY  
 AKLEKINAEQQLQELQELKEKAVASTAETALONTROAKERALEKQNDSE  
 SEGRKLVAEELBEKRAVOASOHEVHLKQKOHBEKIKVLDQIKDLDKES  
 TLENNQREHEEHEKIKLSEOKAMINAMSKIRISORIVELSEANKLANSLFT  
 ORMKROEMTSEIQOKPYLETQAGKLAQRKKEBQLEKISHDSDKRLIELEI  
 RLREVSLEHBEQLEKROLTELQSLQRESQTLALQAPALSSQLQKTELEET  
 TAAEEIQTALTAHDEIQKFTDALNRSCTVITLDEQLNLTEDNALNNQNFYLSK

BASE COUNT 2277 a 2204 c 2220 g 1875 t  
 ORIGIN

Query Match 94.8%; Score 5843.2; DB 9; Length 8576;  
 Best Local Similarity 97.2%; Pred. No. 0;  
 Matches 6036; Conservative 0; Mismatches 3; Indels 171; Gaps 2;

QY 1 ATGTTGAAGTCAAAATATGAGAGCGCGGAACTCTTGGATGCTGCTGCTGAACCAT 60  
 DB 54 ATGTTGAAGTCAAAATATGAGAGCGCGGAACTCTTGGATGCTGCTGCTGAACCAT 113  
 QY 61 GCCAGCGGCGCTCCAGGCTGAATCTGCTTCCAGGAGAACACCCCTTATGACTCAA 120  
 DB 114 GCCAGCGGCGCTCCAGGCTGAATCTGCTTCCAGGAGAACACCCCTTATGACTCAA 113  
 QY 121 CAGCAGATGCTCTCTCTTCCGAGAGGAGATATGATCCCTCTTGTCTCTTGA 180  
 DB 174 CAGCAGATGCTCTCTCTTCCGAGAGGAGATATGATCCCTCTTGTCTCTTGA 233  
 QY 181 GAATGAGTCAAGCGCTGCTGATGAAGATTAACGAGTGAAGCTTGTCCGAGATAT 240  
 DB 234 GAATGAGTCAAGCGCTGCTGATGAAGATTAACGAGTGAAGCTTGTCCGAGATAT 233  
 QY 241 TCCGACACCAATAGCTGATTAACGAGCTTCCGAGAGGAGCTTGAAGTCA 300  
 DB 294 TCCGACACCAATAGCTGATTAACGAGCTTCCGAGAGGAGCTTGAAGTCA 353  
 QY 301 AGCTTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 DB 354 AGCTTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 413  
 QY 361 GACATCTATGCTAAGAAAGTGAAGAGAGAGCTTATGAGCCGAGAGAGTTC 420  
 DB 414 GACATCTATGCTAAGAAAGTGAAGAGAGAGCTTATGAGCCGAGAGAGTTC 473  
 QY 421 TTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
 DB 474 TTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 533  
 QY 481 CAGTATCCCTTTCAGAGCAAAATATCACTTATCTGCTGCTGAGATATGAGCTGAGG 540  
 DB 534 CAGTATCCCTTTCAGAGCAAAATATCACTTATCTGCTGCTGAGATATGAGCTGAGG 593  
 QY 541 GACTTGCTGCTGCTTGAATAGATATGAGAGCAAGTATGAGTAAATCTGATACGTTT 600  
 DB 594 GACTTGCTGCTGCTTGAATAGATATGAGAGCAAGTATGAGTAAATCTGATACGTTT 653  
 QY 601 TACCTAGCTAGCTGATTTTGGCTGCTCAGAGCTTCACTGATGAGATGAGTACGATCGA 660  
 DB 654 TACCTAGCTAGCTGATTTTGGCTGCTCAGAGCTTCACTGATGAGATGAGTACGATCGA 713  
 QY 661 GACATCAAGCTTGAAGACATTTCTGCTGAGCCGACAGAGACATCAAGCTGATGATTT 720  
 DB 714 GACATCAAGCTTGAAGACATTTCTGCTGAGCCGACAGAGACATCAAGCTGATGATTT 773

OY	721	GGAATCGCGCGGAAAATGAATTCGAACAAGATGGGAATGCGCAACTCCGATTGGGACC	780
Db	774	GGATTCGCGCGAAAATGAAATTCGAACAAGATGGGAATGCGCAACTCCGATTGGGACC	833
OY	781	CCAGATTACATGCGCTCCTGAAGTGTGACTGTGATGAACCGGGATGAGAAAAGCACCTAC	840
Db	834	CCAGATTACATGCGCTCCTGAAGTGTGACTGTGATGAACCGGGATGAGAAAAGCACCTAC	893
OY	841	GGCCGTGACCTGTGAATCGGTGGTCACTGGGCGTGAATTGCTCATGAGATGATTTATGGAGA	900
Db	894	GGCCGTGACCTGTGAATCGGTGGTCACTGGGCGTGAATTGCTCATGAGATGATTTATGGAGA	953
OY	901	TCGCCCTTCGCAAGAGGAACTCTGCCGAACCTTCATTAACATATGAAATTTCCAGAGG	960
Db	954	TCGCCCTTCGCAAGAGGAACTCTGCCGAACCTTCATTAACATATGAAATTTCCAGAGG	1013
OY	961	TTTTTGAAATTTCCAGATGACCCCAAGTGAAGCACTTCTTTGATCTGATTCGAATTCAAAGC	1020
Db	1014	TTTTTGAAATTTCCAGATGACCCCAAGTGAAGCACTTCTTTGATCTGATTCGAATTCAAAGC	1073
OY	1021	TTGTGTGCGCGCCAGAAAAGAGACTGAAGTTGAAGGCTTTGTCGCATCCTTCTTC	1080
Db	1074	TTGTGTGCGCGCCAGAAAAGAGACTGAAGTTGAAGGCTTTGTCGCATCCTTCTTC	1133
OY	1081	TCATAAATTGACTGGAACACATTCGTAACTCTCTCCCTCCCTTCGTCCACCTCAAG	1140
Db	1134	TCATAAATTGACTGGAACACATTCGTAACTCTCTCCCTCCCTTCGTCCACCTCAAG	1193
OY	1141	TCTGCATGTACACCTCCCAATTTTATATGAACAGAGAAATTCGTGGGTTCAATCCCT	1200
Db	1194	TCTGCATGTACACCTCCCAATTTTATATGAACAGAGAAATTCGTGGGTTCAATCCCT	1255
OY	1201	CCGTGCCAGCTGAGGCTCCTCAGGCTTCTCGGCTGAAGACTGCCGTTTGTGGGGTTTTCG	1260
Db	1254	CCGTGCCAGCTGAGGCTCCTCAGGCTTCTCGGCTGAAGACTGCCGTTTGTGGGGTTTTCG	1313
OY	1261	TACAGCAAGGCACTGGGGATTTCTTGCTGATCTGACTGTTGTTGTCGGGTCTGGACCTC	1320
Db	1314	TACAGCAAGGCACTGGGGATTTCTTGCTGATCTGACTGTTGTTGTTGTCGGGTCTGGACCTC	1373
OY	1321	CTCGCCAAAGACTAGCTCCATGGAAGAAAGAACTTCTCATCAAAAGCAAGACCTACAAAGC	1380
Db	1374	CTCGCCAAAGACTAGCTCCATGGAAGAAAGAACTTCTCATCAAAAGCAAGACCTACAAAGC	1433
OY	1381	TCTCAGAGCAAGTGTCAACAATGAGAGAGAAATGACCCGGTTATCATCTGAGAGTGTCA	1440
Db	1434	TCTCAGAGCAAGTGTCAACAATGAGAGAGAAATGACCCGGTTATCATCTGAGAGTGTCA	1493
OY	1441	GAGGTGAGGCTGTGCTTAAGTCAAGAGAGGTGAGCTGAAGGCTCTGAGACTCAGAGA	1500
Db	1494	GAGGTGAGGCTGTGCTTAAGTCAAGAGAGGTGAGCTGAAGGCTCTGAGACTCAGAGA	1553
OY	1501	TCCCTCCCTGAGAGCAGGACCTTGCTACCTAACATCAACGAATCAGTAGCTTTAAAGGAAGT	1566
Db	1554	TCCCTCCCTGAGAGCAGGACCTTGCTACCTAACATCAACGAATCAGTAGCTTTAAAGGAAGT	1613
OY	1561	TTGAGCAAGCAACGATGAGAGTGTCCAGAGAGATGAACAAGCACTTGACACTTCTCCAT	1620
Db	1614	TTGAGCAAGCAACGATGAGAGTGTCCAGAGAGATGAACAAGCACTTGACACTTCTCCAT	1673
OY	1621	GATATCAAGAGCAGAGGCGGAGAGTCCAAAGAAATCAAGAGCAGGAACTACAGGCTCA	1680
Db	1674	GATATCAAGAGCAGAGGCGGAGAGTCCAAAGAAATCAAGAGCAGGAACTACAGGCTCA	1733
OY	1681	GTGGAAGAAATGAGGTTGATATGATTAATCATGTTGGAAGAGATCTTGTCTCAGCAAGAGA	1740
Db	1734	GTGGAAGAAATGAGGTTGATATGATTAATCATGTTGGAAGAGATCTTGTCTCAGCAAGAGA	1793
OY	1741	CGAGATGATCTCTACGAATCTGAGGTGAGAGATCTCGGCTTGCTGCTGAGAAATTCAG	1800
Db	1794	CGAGATGATCTCTACGAATCTGAGGTGAGAGATCTCGGCTTGCTGCTGAGAAATTCAG	1853

QY	1801	CGGAAACGCA	CAGAA	TGTCAG	CATPAA	CTGTTGA	GGCTA	AGGATCA	AGGAAC	CTCGAA	1860
Db	1854	CGGAACCGA	CAGAA	TGTCA	GCATPAA	CTGTGA	AGGCTA	AGGATCA	AGGAAG	CTCGAA	1913
QY	1861	GTGGAGATAT	TGCGAA	AACTGG	AGAGATCA	ATGCTG	AGCAG	CA	GCCTCA	AAATTTCA	1920
Db	1914	GTGGAGAAAT	TGCGAA	AACTGG	AGAAATCA	ATCTG	AGCAG	CA	GCCTCA	AAATTTCA	1973
QY	1921	CTCCAGAGAA	AACTGG	AGAAAG	CGTGTAA	AAAGCCG	ACG	AGAGG	CCACCG	AGCTCTG	1980
Db	1974	CTCCAGAGAA	AACTGG	AGAAAG	CGTGTAA	AAAGCCG	ACG	AGAGG	CCACCG	AGCTCTG	2033
QY	1981	AATATCGCC	AGCAAG	AGAGCG	CGAGG	AGGAGG	AGTGTG	AGAACTG	CAAA	CCGAGAG	2040
Db	2034	AATATCGCC	AGCAAG	AGAGCG	CGAGG	AGGAGG	AGTGTG	AGAAAG	CTGCA	AAACCGAGAG	2093
QY	2041	GATTCTTCT	GAAAGCA	TCAGAA	AAGCTGTG	GAAGCTG	AGAA	CGCCG	CAATTTCT	CTG	2100
Db	2094	GATTCTTCT	GAAAGCA	TCAGAA	AAGCTGTG	GAAGCTG	AGAA	CGCCG	CAATTTCT	CTG	2134
QY	2101	GAGAA	CAAGGTAA	AGAGCTAG	AGACATG	AGGCGT	AGAGAA	AAACAG	CTGAA	AGATGAC	2166
Db	2135	-----	-----	-----	-----	-----	-----	-----	-----	-----	2134
QY	2161	ATCCAGA	CAAAATCC	ACAGATCC	ACAGATG	GCATPAA	ATTCTG	AGCTCG	AAAG	2220	
Db	2135	-----	-----	-----	-----	-----	-----	-----	-----	-----	2147
QY	2221	AAACATG	GGGAGG	CCCAAG	CTCA	GGCCAG	ACCTA	PAAGTG	CACTTG	AAACAGAA	2280
Db	2148	AAACATG	GGGAGG	CCCAAG	CTCA	GGCCAG	ACCTA	PAAGTG	CACTTG	AAACAGAA	2207
QY	2281	CAGCACTAT	GAGAAAA	AGTTAA	AGTTGGA	CAATCAG	ATPAA	AGAAAG	CACTG	AGCTGAC	2340
Db	2208	CAGCACTAT	GAGAAAA	AGTTAA	AGTTGGA	CAATCAG	ATPAA	AGAAAG	CACTG	AGCTGAC	2267
QY	2341	AAGAGAC	CACTGG	AGAA	CATGATC	AGAGAC	ACAGAG	AGAGG	CCCAT	TAGAGAGG	2400
Db	2268	AAGAGAC	CACTGG	AGAA	CATGATC	AGAGAC	ACAGAG	AGAGG	CCCAT	TAGAGAGG	2327
QY	2401	ATTCTCA	GCAGAA	CGAAG	CGATAT	CAATGCT	TGGAATTC	CAAGATC	AGATCC	CTGGA	2460
Db	2328	ATTCTCA	GCAGAA	CGAAG	CGATAT	CAATGCT	TGGAATTC	CAAGATC	AGATCC	CTGGA	2387
QY	2461	CAGAGAT	TGTGCA	CTGTCT	GAACCA	ATPAACTTG	CA	CAAAATG	ACAGTCT	TTTTAC	2520
Db	2388	CAGAGAT	TGTGCA	CTGTCT	GAACCA	ATPAACTTG	CA	CAAAATG	ACAGTCT	TTTTAC	2447
QY	2521	CAAGGAA	CAATGA	AGGCGC	CAAGAA	GABATG	TTCTG	AAC	TCAGG	CAACGAA	2580
Db	2448	CAAGGAA	CAATGA	AGGCGC	CAAGAA	GABATG	TTCTG	AAC	TCAGG	CAACGAA	2507
QY	2581	CTGGAGA	CACAGG	CTGGAA	AGTTG	AGAGCC	CAGAA	CTGAA	AACTG	AGAGCA	2640
Db	2508	CTGGAGA	CACAGG	CTGGAA	AGTTG	AGAGCC	CAGAA	CTGAA	AACTG	AGAGCA	2567
QY	2641	AAGATCA	GCAC	CAAGAC	CA	CA	GTGAC	CA	AAGAT	TGGCTG	2700
Db	2568	AAGATCA	GCAC	CAAGAC	CA	CA	GTGAC	CA	AAGAT	TGGCTG	2627
QY	2701	CGGAGG	CTCA	GTCTAG	AGCA	CGAGAG	CA	AAATCTG	AG	CTCAAG	2760
Db	2628	CGGAGG	CTCA	GTCTAG	AGCA	CGAGAG	CA	AAATCTG	AG	CTCAAG	2687
QY	2761	CTTACAG	CTTCTC	CTG	CAGAG	CGGAG	CTCA	CA	AGTTG	CAG	2820
Db	2688	CTTACAG	CTTCTC	CTG	CAGAG	CGGAG	CTCA	CA	AGTTG	CAG	2747
QY	2821	GCCTTGA	AGGCA	CGCTTG	CCAGG	CGA	AAACAG	AGTGTG	AAAGAG	CA	2880
Db	2748	GCCTTGA	AGGCA	CGCTTG	CCAGG	CGA	AAACAG	AGTGTG	AAAGAG	CA	2807
QY	2881	GAAAGAG	AGATCA	AGCA	CTCA	CGGCA	CAT	TAGAG	TGAAT	CA	2940









QY 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGGCTTCATCTGATGGGATACGTGATGCA 660  
 Db 601 TACCTAGCTGAGCTGATTTTGGCTGTTCAAGGCTTCATCTGATGGGATACGTGATGCA 660  
 QY 661 GACATCAAGCCTGAGAACATTTCTGTTGACCGACAGACACATCAAGCTGATGATTTT 720  
 Db 661 GACATCAAGCCTGAGAACATTTCTGTTGACCGACAGACACATCAAGCTGATGATTTT 720  
 QY 721 GGAATCTGCGCGGAAAAATGAAATTCAAACAAGATGATGATGCTCAACTCCGATTTGGAC 780  
 Db 721 GGAATCTGCGCGGAAAAATGAAATTCAAACAAGATGATGATGCTCAACTCCGATTTGGAC 780  
 QY 781 CCAGATTACATGAGCTCTGAAAGTGTGACTGTGATGAAAGGAGATGAAAGGACCTAC 840  
 Db 781 CCAGATTACATGAGCTCTGAAAGTGTGACTGTGATGAAAGGAGATGAAAGGACCTAC 840  
 QY 841 GGCCTTGAGCTGTGACTGTGATGATGAGCGCGATGATGCTGATGATGATGATGATGATG 900  
 Db 841 GGCCTTGAGCTGTGACTGTGATGATGAGCGCGATGATGCTGATGATGATGATGATGATG 900  
 QY 901 TCCGCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 Db 901 TCCGCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 QY 961 TTTTGAATTTTCCAGATGACCCCAAGATGAGAGATGATGATGATGATGATGATGATG 1020  
 Db 961 TTTTGAATTTTCCAGATGACCCCAAGATGAGAGATGATGATGATGATGATGATGATG 1020  
 QY 1021 TTTGTTGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 Db 1021 TTTGTTGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
 QY 1081 TCTTAAATTTGATGAGAACCAATTCGTAATCTCTCTCCCTCCCTGTTGATGATGATG 1140  
 Db 1081 TCTTAAATTTGATGAGAACCAATTCGTAATCTCTCTCCCTCCCTGTTGATGATGATG 1140  
 QY 1141 TCTGAGATGACCTCTCAATTTTGTATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 Db 1141 TCTGAGATGACCTCTCAATTTTGTATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 QY 1201 CCGTGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 Db 1201 CCGTGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 QY 1261 TACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 Db 1261 TACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
 QY 1321 CCGTGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
 Db 1321 CCGTGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
 QY 1381 TCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 Db 1381 TCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
 QY 1441 GAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 Db 1441 GAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 QY 1501 TCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 Db 1501 TCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 QY 1561 TTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 Db 1561 TTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 QY 1621 GATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
 Db 1621 GATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
 QY 1681 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740

Db 1681 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
 QY 1741 CGAGATGATCTTACAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800  
 Db 1741 CGAGATGATCTTACAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800  
 QY 1801 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
 Db 1801 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860  
 QY 1861 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
 Db 1861 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920  
 QY 1921 CTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
 Db 1921 CTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
 QY 1981 AATATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
 Db 1981 AATATCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
 QY 2041 GATTCCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100  
 Db 2041 GATTCCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100  
 QY 2101 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
 Db 2101 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160  
 QY 2161 ATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
 Db 2161 ATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
 QY 2221 AAACATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
 Db 2221 AAACATGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
 QY 2281 CAGCATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
 Db 2281 CAGCATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
 QY 2341 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400  
 Db 2341 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400  
 QY 2401 ATTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460  
 Db 2401 ATTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460  
 QY 2461 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520  
 Db 2461 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520  
 QY 2521 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580  
 Db 2521 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580  
 QY 2581 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640  
 Db 2581 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640  
 QY 2641 AAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700  
 Db 2641 AAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700  
 QY 2701 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760  
 Db 2701 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760  
 QY 2761 CTACAGCTCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820

Db 2761 CTAAAGCTCTCCGCGAGAGCGGAGTCAAGTTGACAGCCCTGACAGGCTGCACCGGCG 2820  
Qy 2821 GCCCTGAGAGCCAGCTTCCGCGAGCGGAGACAGAGCTGAGAGAGCCACAGAGAGCT 2880  
Db 2821 GCCCTGAGAGCCAGCTTCCGCGAGCGGAGAGAGAGCTGAGAGAGCCACAGAGAGCT 2880  
Qy 2881 GAAGAGAGAGATCCAGGAGCTCAGGCGCATAGAGATGAAATCCAGGCGAAATTTGATGCT 2940  
Db 2881 GAAGAGAGAGATCCAGGAGCTCAGGCGCATAGAGATGAAATCCAGGCGAAATTTGATGCT 2940  
Qy 2941 CTTGTAACAGCTGTAAGTATGATCAAGACCTGAGAGAGAGCTAAACAGCTGACGAG 3000  
Db 2941 CTTGTAACAGCTGTAAGTATGATCAAGACCTGAGAGAGAGCTAAACAGCTGACGAG 3000  
Qy 3001 GACAAAGCTGAACTCAACCAACCAAACTTCTAATTGCGCAAAATACTGATGAGGCTTCT 3060  
Db 3001 GACAAAGCTGAACTCAACCAACCAAACTTCTAATTGCGCAAAATACTGATGAGGCTTCT 3060  
Qy 3061 GGCAGCAACGACGAGATTTGTACAACTGCGAAGTGAAGTGAACCATCTCCGCGGAGATC 3120  
Db 3061 GGCAGCAACGACGAGATTTGTACAACTGCGAAGTGAAGTGAACCATCTCCGCGGAGATC 3120  
Qy 3121 ACAGAAAGAGAGATGAGCTTACAGCGAAGAGAAAGATGAGAGCTCTGAAGAGCAG 3180  
Db 3121 ACAGAAAGAGAGATGAGCTTACAGCGAAGAGAAAGATGAGAGCTCTGAAGAGCAG 3180  
Qy 3181 TGCAACATGCTGAGAGAAACAGATCATGATTTGAGAGCCCTAAACAGATGAGCTAGAA 3240  
Db 3181 TGCAACATGCTGAGAGAAACAGATCATGATTTGAGAGCCCTAAACAGATGAGCTAGAA 3240  
Qy 3241 AAAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300  
Db 3241 AAAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300  
Qy 3301 TGTGAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360  
Db 3301 TGTGAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360  
Qy 3361 GATGAGAGAGATCAACGAGTCTCCGAGAGTGTGAGAGTGGCAATGAAAGAGAGAGAG 3420  
Db 3361 GATGAGAGAGATCAACGAGTCTCCGAGAGTGTGAGAGTGGCAATGAAAGAGAGAGAG 3420  
Qy 3421 GAGATGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480  
Db 3421 GAGATGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480  
Qy 3481 GACAGCTCAATGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
Db 3481 GACAGCTCAATGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540  
Qy 3541 CAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600  
Db 3541 CAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600  
Qy 3601 TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
Db 3601 TTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
Qy 3661 GAACTGTAAGTGGGCTGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720  
Db 3661 GAACTGTAAGTGGGCTGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720  
Qy 3721 GAAAGATTCAGGTTCTCTATCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780  
Db 3721 GAAAGATTCAGGTTCTCTATCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780  
Qy 3781 CAAGCAAACTCAATGATTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840  
Db 3781 CAAGCAAACTCAATGATTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840  
Qy 3841 GTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900  
Db 3841 GTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900

Qy 3901 GAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960  
Db 3901 GAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960  
Qy 3961 GCCCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020  
Db 3961 GCCCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020  
Qy 4021 ATGCGCATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080  
Db 4021 ATGCGCATGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080  
Qy 4081 GCGCGCATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140  
Db 4081 GCGCGCATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140  
Qy 4141 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200  
Db 4141 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200  
Qy 4201 ACAAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260  
Db 4201 ACAAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260  
Qy 4261 GAATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320  
Db 4261 GAATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320  
Qy 4321 CCGTGAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4380  
Db 4321 CCGTGAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4380  
Qy 4381 GGTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4440  
Db 4381 GGTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4440  
Qy 4441 AGAATTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4500  
Db 4441 AGAATTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4500  
Qy 4501 AAAGTCCATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560  
Db 4501 AAAGTCCATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560  
Qy 4561 GAGCTGAGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620  
Db 4561 GAGCTGAGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4620  
Qy 4621 GCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
Db 4621 GCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
Qy 4681 ACTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4740  
Db 4681 ACTGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4740  
Qy 4741 TGGGTCACCGCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800  
Db 4741 TGGGTCACCGCTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800  
Qy 4801 GCTGATGCTAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
Db 4801 GCTGATGCTAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
Qy 4861 ATGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920  
Db 4861 ATGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920  
Qy 4921 TAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4980  
Db 4921 TAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4980

QY 4981 CAAATTTATTTATTCAGGAGCTGGAAGCTTATCTATATGAGAGAGAGCGGCA 5040  
 Db 4981 CAAATTTATTTATTCAGGAGCTGGAAGCTTATCTATATGAGAGAGAGCGGCA 5040  
 QY 5041 CTGTGCTTGTGAGCTGGAAGAGTGAAGAGTCCCTGCGCCAGTCCACCTGCTGCC 5100  
 Db 5041 CTGTGCTTGTGAGCTGGAAGAGTGAAGAGTCCCTGCGCCAGTCCACCTGCTGCC 5100  
 QY 5101 CAGCCGAGATCTCAACCAATTTTGAAGCTGTCAGAGGCTGCTGCTGCTGCTGCA 5160  
 Db 5101 CAGCCGAGATCTCAACCAATTTTGAAGCTGTCAGAGGCTGCTGCTGCTGCTGCA 5160  
 QY 5161 GGGAAATTTGAGAGCGGCTTGTGATCTGTGAGAGCTGCTGCTGCTGCTGCTGCTG 5220  
 Db 5161 GGGAAATTTGAGAGCGGCTTGTGATCTGTGAGAGCTGCTGCTGCTGCTGCTGCTG 5220  
 QY 5221 CGCTAAGCAAAACCTCAGCAAAATCTGATCCGGAAGAGATGAGAGCTCAGAGCC 5280  
 Db 5221 CGCTAAGCAAAACCTCAGCAAAATCTGATCCGGAAGAGATGAGAGCTCAGAGCC 5280  
 QY 5281 TGGAGCTGATTCATCTTCAACCAATTACATATCTTATTTGAAACCAATTAATTTCA 5340  
 Db 5281 TGGAGCTGATTCATCTTCAACCAATTACATATCTTATTTGAAACCAATTAATTTCA 5340  
 QY 5341 ATGACATGAGAGAGTACAGCTGAGAGATTCCTGATAGATGATGACCATTCCTTGGCA 5400  
 Db 5341 ATGACATGAGAGAGTACAGCTGAGAGATTCCTGATAGATGATGACCATTCCTTGGCA 5400  
 QY 5401 CCTGCTGTGTTTCCGCTCTTTCACACAGCTTCCCTGCTCAATGCTGAGGTGAACAGC 5460  
 Db 5401 CCTGCTGTGTTTCCGCTCTTTCACACAGCTTCCCTGCTCAATGCTGAGGTGAACAGC 5460  
 QY 5461 GCGAGGAGAGAGAGAGTACTGCTGTGTTTCCAGATTTGAGAGTGTGCTGATGCT 5520  
 Db 5461 GCGAGGAGAGAGAGAGTACTGCTGTGTTTCCAGATTTGAGAGTGTGCTGATGCT 5520  
 QY 5521 TACGGAAGAGTACGCGACAGAGAGTCTCAAGTGAAGTCCCTTGGCTTGGCC 5580  
 Db 5521 TACGGAAGAGTACGCGACAGAGAGTCTCAAGTGAAGTCCCTTGGCTTGGCC 5580  
 QY 5581 TACGGAAGAGTACGCGACAGAGAGTCTCAAGTGAAGTCCCTTGGCTTGGCC 5640  
 Db 5581 TACGGAAGAGTACGCGACAGAGAGTCTCAAGTGAAGTCCCTTGGCTTGGCC 5640  
 QY 5641 GCAAGCTCTCTCAGAGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5700  
 Db 5641 GCAAGCTCTCTCAGAGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5700  
 QY 5701 CTGGGCTCTGCTCAGAGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5760  
 Db 5701 CTGGGCTCTGCTCAGAGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5760  
 QY 5761 AGGGTCAATTTGCTGAGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5820  
 Db 5761 AGGGTCAATTTGCTGAGAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5820  
 QY 5821 CCGTCCAGCTCCCGAG 5837  
 Db 5821 CCGTCCAGCTCCCGAG 5837

RESULT 11  
 AX429514 6159 bp DNA linear PAT 21-JUN-2002  
 LOCUS AX429514  
 DEFINITION Sequence 3 from Patent WO0234896.  
 ACCESSION AX429514  
 VERSION AX429514.1 GI:21540792  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1

AUTHORS Kapeller-Liebermann, R.  
 TITLE 13245, a novel human myotonic dystrophy type protein kinase and  
 JOURNAL uses therefor  
 Patent: WO 0234896-A 3 02-MAY-2002;  
 FEATURES MILLENNIUM PHARM INC (US)  
 source location/Qualifiers  
 1. 6159  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 1745 a 1528 c 1679 g 1207 t  
 ORIGIN  
 Query Match 91.8%; Score 5661.4; DB 6; Length 6159;  
 Best Local Similarity 95.8%; Pred. No. 0;  
 Matches 5944; Conservative 0; Mismatches 16; Indels 243; Gaps 3;  
 QY 1 ATGTTGAAGTCAATATATGAGAGCGGAGATCCTTTGATGCTGTGCTGGAACCAT 60  
 Db 1 ATGTTGAAGTCAATATATGAGAGCGGAGATCCTTTGATGCTGTGCTGGAACCAT 60  
 QY 61 GCGAGCGGAGCTTCAAGCTGAATCTGTTCTCCAGGGAACCAACCTTTATGACTCA 120  
 Db 61 GCGAGCGGAGCTTCAAGCTGAATCTGTTCTCCAGGGAACCAACCTTTATGACTCA 120  
 QY 121 CAGGAGATGCTCTCTCTTCCGAGAGAGATTTAGATGCTGCTGCTGCTGCTGCTG 180  
 Db 121 CAGGAGATGCTCTCTCTTCCGAGAGAGATTTAGATGCTGCTGCTGCTGCTGCTG 180  
 QY 181 GAATGAGTCAAGCTGCTGATGATGATTAAGACGTGAGCACTTTGCTCGGAAAT 240  
 Db 181 GAATGAGTCAAGCTGCTGATGATGATTAAGACGTGAGCACTTTGCTCGGAAAT 240  
 QY 241 TCCGACACATAGTGAAGTACAGAGCTCCAGCTTCCGCAAGAGCTTGAAGTCA 300  
 Db 241 TCCGACACATAGTGAAGTACAGAGCTCCAGCTTCCGCAAGAGCTTGAAGTCA 300  
 QY 301 AGCTCTGAGGTTGTGCTCACTTGTGAGAGTGAAGTGAAGAGAGAGAGAGAGAG 360  
 Db 301 AGCTCTGAGGTTGTGCTCACTTGTGAGAGTGAAGTGAAGAGAGAGAGAGAGAG 360  
 QY 361 GACATCTATGCTATGAAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
 Db 361 GACATCTATGCTATGAAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
 QY 421 TTTTGAAGAGAGAGAGAGAGATTTATCTGAGAGCAAGAGAGAGAGAGAGAG 480  
 Db 421 TTTTGAAGAGAGAGAGAGAGATTTATCTGAGAGCAAGAGAGAGAGAGAGAG 480  
 QY 481 CAGATAGCTTTCAGAGCAAAATACCTTTATCTGATGAGAGATTCAGCTGAGAG 540  
 Db 481 CAGATAGCTTTCAGAGCAAAATACCTTTATCTGATGAGAGATTCAGCTGAGAG 540  
 QY 541 GACTTGTGCTGCTTGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 Db 541 GACTTGTGCTGCTTGAATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 QY 601 TACCTAGTGAAGTATTTGCTGCTTCAAGCTTCACTGATGAGAGATTCAGCTG 660  
 Db 601 TACCTAGTGAAGTATTTGCTGCTTCAAGCTTCACTGATGAGAGATTCAGCTG 660  
 QY 661 GACATCAAGCTTGAAGCAATTTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 Db 661 GACATCAAGCTTGAAGCAATTTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAG 720  
 QY 721 GATCTGCGGAGAGAGAGAGATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 Db 721 GATCTGCGGAGAGAGAGAGATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 QY 781 CCAATTTCAATGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
 Db 781 CCAATTTCAATGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840













Db 2851 GAAGAGAGATCCAGGCACTTCACGGCACAATAGAGATGAATCCAGCGCAATTGTGCT 2910  
QY 2941 CTTGTAAACAGCTGTACTGTATATCAAGACTGGAGAGAGAGCTAAACCGAGCTGACCGAG 3000  
Db 2911 CTTGTAAACAGCTGTACTGTATATCAAGACTGGAGAGAGAGCTAAACCGAGCTGACCGAG 2970  
QY 3001 GACACAGCTGAATCTCAACCAACCAAACTTCTACTTGTCCAAACCACTGATGAGGCTTCT 3060  
Db 2971 GACACAGCTGAATCTCAACCAACCAAACTTCTACTTGTCCAAACCACTGATGAGGCTTCT 3030  
QY 3061 GGGCCCAACGACGAGATTTGTAACAATGCGAAAGTGAATGACCAATCTCCCGCGAGATC 3120  
Db 3031 GGGCCCAACGACGAGATTTGTAACAATGCGAAAGTGAATGACCAATCTCCCGCGAGATC 3090  
QY 3121 ACCGACGAGAGATGCACTTACAGCGAGAGAGAGCTCTGGGTATGAGAAATCCGAGTTGAG 3180  
Db 3091 ACCGACGAGAGATGCACTTACAGCGAGAGAGAGCTCTGGGTATGAGAAATCCGAGTTGAG 3150  
QY 3181 TGACCAATGCTGAGAGAAACAGGTCATGATTTGAGAGGCTTAAACGATGAGCTGTAGAA 3240  
Db 3151 TGACCAATGCTGAGAGAAACAGGTCATGATTTGAGAGGCTTAAACGATGAGCTGTAGAA 3210  
QY 3241 AAAGAGGAGAGTGGAGAGGCTGAGAGAGGCTCTGGGTATGAGAAATCCGAGTTGAG 3300  
Db 3211 AAAGAGGAGAGTGGAGAGGCTGAGAGAGGCTCTGGGTATGAGAAATCCGAGTTGAG 3270  
QY 3301 TGTGGGTTTGAAGCTTGAAGAGATCTGAGACACGAGAAACGAGAGAGAGAGCC 3360  
Db 3271 TGTGGGTTTGAAGCTTGAAGAGATCTGAGACACGAGAAACGAGAGAGAGAGCC 3330  
QY 3361 GATAGAGGATCAACCGAGCTCCGCAAGTGTGAGAGCTGAGAGAGAGAGAGAGAGCT 3420  
Db 3331 GATAGAGGATCAACCGAGCTCCGCAAGTGTGAGAGCTGAGAGAGAGAGAGAGAGCT 3390  
QY 3421 GAGATTCGCTGCTGACAGAGAGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGCT 3480  
Db 3391 GAGATTCGCTGCTGACAGAGAGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGCT 3450  
QY 3481 GACAGCTCAATGACCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 3540  
Db 3451 GACAGCTCAATGACCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 3510  
QY 3541 CAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600  
Db 3511 CAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3570  
QY 3601 TTACAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
Db 3571 TTACAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3630  
QY 3661 GAAGCTCTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720  
Db 3631 GAAGCTCTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3690  
QY 3721 GAAAGCAATTCAGGTTCTTATTTCTATGAAAGGTGAAATGAGAGAGAGAGAGAGAG 3780  
Db 3691 GAAAGCAATTCAGGTTCTTATTTCTATGAAAGGTGAAATGAGAGAGAGAGAGAGAG 3750  
QY 3781 CAAGCAATTCATGATTTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3838  
Db 3751 CAAGCAATTCATGATTTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3810  
QY 3839 -----AGTTCTCTGACGATC 3855  
Db 3811 GGTATTATTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3870  
QY 3856 AATGAGCTGAAGCTGGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3915  
Db 3871 AATGAGCTGAAGCTGGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3930  
QY 3916 CTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3975  
Db 3931 CTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3990

QY 3976 AAGGACCAACCCACACCCATTCACGCGACGACCCGAGAGAGAGAGAGAGAGAGAGAG 4035  
Db 3991 AAGGACCAACCCACACCCATTCACGCGACGACCCGAGAGAGAGAGAGAGAGAGAGAG 4050  
QY 4036 ATGTCGAGTCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4095  
Db 4051 ATGTCGAGTCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4110  
QY 4096 CGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4155  
Db 4111 CGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4170  
QY 4156 CACAAATATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4215  
Db 4171 CACAAATATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4230  
QY 4216 TGTCTGAGATCCGTCACCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4275  
Db 4231 TGTCTGAGATCCGTCACCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4290  
QY 4276 TGTCTGAGATCCGTCACCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4335  
Db 4291 TGTCTGAGATCCGTCACCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4350  
QY 4336 ACACACTTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4395  
Db 4351 ACACACTTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4410  
QY 4396 GAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4455  
Db 4411 GAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4470  
QY 4456 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4515  
Db 4471 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4530  
QY 4516 GACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4575  
Db 4531 GACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4590  
QY 4576 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4635  
Db 4591 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4650  
QY 4636 GCAAGATGTCATATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4695  
Db 4651 GCA----- 4653  
QY 4696 AGAACCCTCTACTTGTAGCTTCCAGCTTCCCTGACAAACAGAGAGAGAGAGAGAGAG 4755  
Db 4654 ----- 4653  
QY 4756 GAATCAGTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4815  
Db 4654 -----GAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
QY 4816 CTTGAGAAATCTCTGCTGAGAAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4875  
Db 4681 CTTGAGAAATCTCTGCTGAGAAATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4740  
QY 4876 CCGTTCAAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4935  
Db 4741 CCGTTCAAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4800  
QY 4936 TTGAAAAATCTCCATACCATGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4995  
Db 4801 TTGAAAAATCTCCATACCATGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
QY 4996 AAGGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5055  
Db 4861 AAGGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4920

QY 5056 GTGAAAGAAAGTGAAGAGTCCCTGCGCCGCTCCACCTTGCCTCCGACGCCGACATCTCA 5115  
 DB 4921 GTGAAAGAAAGTGAAGAGTCCCTGCGCCGCTCCACCTTGCCTCCGACGCCGACATCTCA 4980  
 QY 5116 CCCAATTTTGAAGCTGTCAAGGGCTGCCATTGTTGGGGGCAAGAGATTGAGAAC 5175  
 DB 4981 CCCAATTTTGAAGCTGTCAAGGGCTGCCATTGTTGGGGGCAAGAGATTGAGAAC 5040  
 QY 5176 GGGCTGTGATCTGTGACAGCCATGCCCAGCAAGTGTCTTCTCCGCTACACGAAAC 5235  
 DB 5041 GGGCTGTGATCTGTGACAGCCATGCCCAGCAAGTGTCTTCTCCGCTACACGAAAC 5100  
 QY 5236 CTGAGCAATATCTGCAATCCGGAAGAGATAGAGACCTCGAGACCCCTGAGCTGTTCAC 5295  
 DB 5101 CTGAGCAATATCTGCAATCCGGAAGAGATAGAGACCTCGAGACCCCTGAGCTGTTCAC 5160  
 QY 5296 TTGACCAATTTACAGTATCTCTCATTTGGAACCAATTAATTTTACGAAATGACATGAGAG 5355  
 DB 5161 TTGACCAATTTACAGTATCTCTCATTTGGAACCAATTAATTTTACGAAATGACATGAGAG 5220  
 QY 5356 TACAGCTGTGAGAAATTTCTGTGATAGAAATGACATTTCTTGGACCTGTGTGTTC 5415  
 DB 5221 TACAGCTGTGAGAAATTTCTGTGATAGAAATGACATTTCTTGGACCTGTGTGTTC 5280  
 QY 5416 GCCTCTTCCAAAGCTTCCCTGTCTCAATCTGCAAGTGAAGAGGCAAGGCAAGGAGAG 5475  
 DB 5281 GCCTCTTCCAAAGCTTCCCTGTCTCAATCTGCAAGTGAAGAGGCAAGGCAAGGAGAGAG 5340  
 QY 5476 GAGTACTTGTGTGTCTTCCAGAAATTTGAGAGTGTGTGTGTGTCTTACGAAAGAGTAC 5535  
 DB 5341 GAGTACTTGTGTGTCTTCCAGAAATTTGAGAGTGTGTGTGTGTCTTACGAAAGAGTAC 5400  
 QY 5536 CGACAGAGATCTCAAGTGAAGTCTGCTTACCTTGGCTTGGCTTACGAGAGACCTAT 5595  
 DB 5401 CGACAGAGATCTCAAGTGAAGTCTGCTTACCTTGGCTTGGCTTACGAGAGACCTAT 5460  
 QY 5596 CTGTGTGTGTGTCTTCACTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 5655  
 DB 5461 CTGTGTGTGTGTCTTCACTTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 5520  
 QY 5656 GGGAGCCCTGCGCGAGAGTACCTGAGATCCGGAACCCGCGTACCTGCGCGCTGCGCATT 5715  
 DB 5521 GGGAGCCCTGCGCGAGAGTACCTGAGATCCGGAACCCGCGTACCTGCGCGCTGCGCATT 5580  
 QY 5716 TCTCTGAGAGCGATTTTACTTGGCTCTCTATACAGAGATTAATTAAGGGTCACTTGTGC 5775  
 DB 5581 TCTCTGAGAGCGATTTTACTTGGCTCTCTATACAGAGATTAATTAAGGGTCACTTGTGC 5640  
 QY 5776 AAGGAAACCTGCTGAGAGAGTCCGCGACCTGAAACCAACCGGGGCGCGCTGCGCATT 5835  
 DB 5641 AAGGAAACCTGCTGAGAGAGTCCGCGACCTGAAACCAACCGGGGCGCGCTGCGCATT 5700  
 QY 5836 AGGAGCCCAACAGGAGAGGCGCCACACCTGTAACAGACATCAACAGGCGGTGAGC 5895  
 DB 5701 AGGAGCCCAACAGGAGAGGCGCCACACCTGTAACAGACATCAACAGGCGGTGAGC 5760  
 QY 5896 TCCAGCCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5955  
 DB 5761 TCCAGCCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5820  
 QY 5956 TACCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6015  
 DB 5821 TACCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5880  
 QY 6016 GAGAAATTTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6075  
 DB 5881 GAGAAATTTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5940  
 QY 6076 GAGAGAGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6135  
 DB 5941 GAGAGAGAGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6000  
 QY 6136 AACCAAGTCTGGAGCCAGTCTTC 6158

Db 6001 AACAGGAAAGAGGAGAGTGC 6023  
 RESULT 13  
 AF086824  
 LOCUS  
 DEFINITION  
 Mus musculus rho/rac-interacting citron kinase (Crik) mRNA,  
 complete cds.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 (bases 1 to 6954)  
 Di Cunto, F., Calautti, E., Hsiao, J., Ong, L., Topley, G., Turco, E. and  
 Dotto, G. P.  
 Cytion rho-interacting kinase, a novel tissue-specific ser/thr  
 kinase encompassing the Rho-Rac-binding protein Cytion  
 J. Biol. Chem. 273 (45), 29706-29711 (1998)  
 9792683  
 2 (bases 1 to 6954)  
 Di Cunto, F., Calautti, E., Hsiao, J., Ong, L., Topley, G., Turco, E. and  
 Dotto, G. P.  
 Direct Submission  
 Submitted (25-AUG-1998) Biology, University of Torino, Via Santena  
 5 bis, Torino, TO 10126, Italy  
 Location/Qualifiers  
 1..6954  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /tissue\_type="spleen"  
 1..6954  
 /gene="Crik"  
 772..6939  
 /gene="Crik"  
 /function="protein kinase"  
 /note="Crik; rho/rac binding protein; alternatively  
 spliced form of citron encoded by GenBank Accession Number  
 AF039218; similar to Mus musculus citron-K encoded by  
 GenBank Accession Number AF070066; similar to Rattus  
 norvegicus citron-K encoded by GenBank Accession Number  
 AF070065"  
 /codon\_start=1  
 /product="rho/rac-interacting citron kinase"  
 /protein\_id="AAC72823.1"  
 /db\_xref="GI:3599509"  
 /translation="MLKFKYGVNPPPEASBPASRLNLFPGKPLMTQOMS  
 ALRREGMLDLFLFECSQPALMKKKHVSFVOKSDPTLAEIRLQSPARDEFVSL  
 VGGHFAEVQVLEKATQDYAMKIMKKKALLAQOVFFEEERNLISSTPMIOL  
 QVAFQDKNNILVMEYOPGGDFSLINRYEDOLDESMIOFYLAELILAVSHOMGYV  
 HDIKPENLILDRTEGIKYDRBSAKNNNSYADALPTGTPDYNAPEVLTVMNER  
 GTTGDCDMHVSQVAVEMVYGTPTPTGSAFTNNIMFQRLFPDPKYSSEL  
 DILQSLICVQKERLKEGLCHPFPARTVMNIRNSPPVPTLKSDDTISNDEBK  
 NWSAFLICVPAELAFSGELPVGFSYKALGYLGRSSVVSLSDPKVSMEKUL  
 LIRKSELSDSDCHMEQDMTELRHREVEVALVLSQKEVELKASRTQSLSEDLAT  
 YIKESLIRKSLSDARMEVSDQDKALDILDRBSRKLQELKEQVQAYEEMLM  
 MNLREDIVASRRSLYSELRESLAEFPKPKRRCQHLMLKAKDKGQPVGVGS  
 KLEKINABOIKOIRQKLEKAVASTATLQNTIRQKBAEIRLEKINRNVGS  
 EGIKKQVLEBERSHLENKVKLEMERENKODITQKSDIOMDKILIEBK  
 HREAOVSQHLVHLKQKQHYEERIKVLNDQIKLADKESLENMROEERABEG  
 KIISQKAMINADSKIRSLBORIVELSNKILAAQNSLFTQENMKAQOEMISELQ  
 KFLYLETQAGLEAKONRKELEKISHOHSDKSLLESLREVSLEHBEQKELK  
 RQVLETOISHOERSQUTALQABALBSOLRQKTELETTAEAEELIQAALTAHDE  
 IORPDLNRNSCVITLDEOLNOLTEDEALINNOYFSLDEKDAAGANDEIVORS  
 EVDHLEPRTTEREMQLSQKQMEALKTCYHLEEVYLDLEALNDELAKEXQMEWR  
 SYLDEKSGEPCREVLQRMQLTBKOSRAPADRTESRQVVELAVKRAKELIALQO  
 ALKEQKLKABESDQKNDLEKHAMLEWARSLOQLTERELKQRLKQKLDQOM  
 DLQKNHIFRLTQGLQALRADLILKTERSDLEVOLENIOVLVSHEKVMKEGTISQTK

LIDFLAKMDPAKKKKVPLQVNEKLALEREKARCALEALQTRLEISAREEA  
 HKRATPHRPHSTPAIRKQOLMSALVRSSEHOPSMASLLAPSSRRKSSSTPEERSR  
 LKERMHNI PHRNVLGNKRAKCAVLTIVHQAQSLCCLQVWCHPKSTCLPAT  
 CGLPAAYATHEAFCDKXNPSBGLQSKBPSLHLEGMKPRANKKQOQMDVKYI  
 VLEBSKVLIVDNEBARQORPVEBELCLPDDVSIHGAVSELANKADVDYILK  
 MESPHTCWPGRITLYLAPSPDKORWTALESYAVAGVREKADPAKLYSL  
 KLEGGDRLDNCTLPSPDOVLVGTREGLYALNVLKNSLTHI BGIAGFOLYI IUDLE  
 KLIMJGEBRALCLDVKKVKSIALOSHI PAODVSPNIFKAVCHLPAKIKNSL  
 CTCAAMPKSVILIRYNDNSKTCIREITSEPCSCIFHTNLSILGNKFEIDMKY  
 YLIDFLKNDHSLAPVAPSSNSFPVSI VOANSQOREYLLCFHGRVFDVGR  
 RSTDLKWSRLFLAPAREPYLFVTHFSLVLEIQAASISGAPARVLEIPRVL  
 GPAISSGAIYLAASSYODKLRVLCNGNLVKESTGEORVPSSTRSSPMKRGPTNEH  
 ITRVASSPAPRPSHPRPSCTPHRYRREGRTELRDRKSGRPLERBSKGRMLST  
 RRRSPGRFEDBSRRLPAGAVRPLPSQVKNVMDQSSV"

BASE COUNT 1779 a 1881 c 1972 g 1322 t  
 ORIGIN

Query Match 82.8%; Score 5105.2; DB 10; Length 6954;  
 Best Local Similarity 89.6%; Pred. No. 0;  
 Matches 5533; Conservative 0; Mismatches 628; Indels 11; Gaps 4;

Qy	1	ATGTGAAGTTCAAAATATGAGGCGGAATCCTTTGAGTGTGCTGCTCAACCAT	60
Db	772	ATGTGAAGTTCAAGTATGTTGTCGGAACCGCCGAGGCAAGCTTCGAGCCCAT	831
Qy	61	GCCAGCCGGGCTCCAGGCTGAATCTGTTCTTCAGAGGAAACACCTTTATGACTCA	120
Db	832	GCCAGTGGGCTCCAGGCTAAATCTTCTTCAGAGGAAACCGCCCTCATGACTCA	891
Qy	121	CAGCAGATGTCCTCTTCCCGAAGAGGATATTGATGCTGCTCTTCTTCTTTGAA	180
Db	892	CAGCAGATGTCCTCTTCCCGAAGAGGATATTGATGCTGCTCTTCTTCTTTGAA	951
Qy	181	GAATGAGTGAAGCTGTCTGATGAAGATTAAAGCAGTGAACAATTGTCCGAAAGTAT	240
Db	952	GAGTGAAGCTGAAGCTGTCTGATGAAGATTAAAGCAGTGAACAATTGTCCGAAAGTAT	1011
Qy	241	TCCGACACCACTAGTGAATGACAGAGCTCCAGCTTCGCGAAGAGACTTCGAGTCA	300
Db	1012	TCCGACACCACTAGTGAATGACAGAGCTCCAGCTTCGCGAAGAGACTTCGAGTCA	1071
Qy	301	AGTCTGTAAGTGTGTCTCACTTGTGTCAGTGTGAAGTGTAAAGAGAAACCAACCGG	360
Db	1072	AGCTTGTGGGCTGTGTCTCACTTGTGTCAGTGTGAAGTGTAAAGAGAAACCGG	1131
Qy	361	GACATCTATCTTATGAAAGTGAAGAGAGGCTTTATTGACCCAGAGACAGGTTTCA	420
Db	1132	GACATCTATCTTATGAAAGTGAAGAGAGGCTTTATTGACCCAGAGACAGGTTTCA	1191
Qy	421	TTTTTGAAGAGAGGAGCAATATTATCTGAAAGCAAGCCCGTGAATCCCAATT	480
Db	1192	TTTTTGAAGAGAGGAGCAATATTATCTGAAAGCAAGCCCGTGAATCCCAATT	1251
Qy	481	CAGTATGCTTTTCAAGGCAAAATCACTTTATCTGTCTATGATGAATCAAGCTTGAAG	540
Db	1252	CAGTATGCTTTTCAAGGCAAAATCACTTTATCTGTCTATGATGAATCAAGCTTGAAG	1311
Qy	541	GACTTGTCTCATCTTTGAATAGATATGAGGCAATGATGATGAATCAAGCTTGAAG	600
Db	1312	GACTTGTCTCATCTTTGAATAGATATGAGGCAATGATGATGAATCAAGCTTGAAG	1371
Qy	601	TACCTAGCTGAGCTGATTTGCTGTTCACAGGCTTCATCTGATGGGATACGTGATCA	660
Db	1372	TACCTAGCTGAGCTGATTTGCTGTTCACAGGCTTCATCTGATGGGATACGTGATCA	1431
Qy	661	GACATCAAGCTTGAAGCAATCTGTTGACCGGACAGGACACATCAAGCTTGAAGTTT	720
Db	1432	GACATCAAGCTTGAAGCAATCTGTTGACCGGACAGGACACATCAAGCTTGAAGTTT	1491
Qy	721	GGAATCTGCGGAAATGAATTAACAAGATGTAATGCAAACTCCGATTTGGAGCC	780
Db	1492	GGAATCTGCGGAAATGAATTAACAAGATGTAATGCAAACTCCGATTTGGAGCC	1548

Qy	781	CCAGATTACATGCTCTCTGAAGTCTGACTGTGATGAATGAGGAGATGGAAGGACACTAC	840
Db	1549	CCGATATTACATGCTCTCTGAAGTCTGACTGTGATGAATGAGGAGATGGAAGGACACTAC	1608
Qy	841	GAGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	900
Db	1609	GAGCTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1668
Qy	901	TCCCCCTTCCGAGAGGAACTTCTGCGAGAACTTCAATPACTTATGAATTTCCAGCG	960
Db	1669	ACCCCATTTCAAGAGGAACTTCCGCGAGAACTTCAATPACTTATGAATTTCCAGCG	1728
Qy	961	TTTTTGAATTTCCAGATGACCCCAAGTGAAGGAGTCACTTTGATCTGATTTCAAGC	1020
Db	1729	TTTTTGAATTTCCAGATGACCCCAAGTGAAGGAGTCACTTTGATCTGATTTCAAGC	1788
Qy	1021	TTGTTGTGCGGCGAGAAAGAGACTGAAGTTGAAGTCTTGTCTGCAATCTTCTTC	1080
Db	1789	CTGCTGTGTGCGGCGAGAAAGAGACTGAAGTTGAAGTCTTGTCTGCAATCTTCTTC	1848
Qy	1081	TCTTAAATGACAGGAAACATTTGTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1140
Db	1849	GCCAGAACGAGCTGGAACAACTCTGTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1908
Qy	1141	TCTGACGATGACACTTCAATTTTGAATGAACAGAGAAATTCGTGGG-TTTCATCTTC	1199
Db	1909	TCTGACGATGACACTTCAATTTTGAATGAACAGAGAAATTCGTGGGCTTTCATCTTC	1968
Qy	1200	TCCGTGCACTGAGACCCCTCAGGCTTCCTGGGTGAAGAACTGCGCTTGTGGGTTTC	1259
Db	1969	TGTTGTGCACTGAGACCCCTCAGGCTTCCTGGGTGAAGAACTGCGCTTGTGGGTTTC	2027
Qy	1260	GTAACAGAAAGCACTGGGATCTTGTGATGATGATGATGATGATGATGATGATGATG	1319
Db	2028	GTAACAGAAAGCACTGGGATCTTGTGATGATGATGATGATGATGATGATGATGATG	2087
Qy	1320	CCCTGCCAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1379
Db	2088	CCCTGCCAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	2147
Qy	1380	CTCTCAGAGCAAGTGTCACAAGATGAGAGCAAGAAATGACCCGTTTACATCGAGTGT	1439
Db	2148	CTCTCAGAGCAAGTGTCACAAGATGAGAGCAAGAAATGACCCGTTTACATCGAGTGT	2207
Qy	1440	AGAGTGAAGGCTGTGCTTATGTCAGAGAGGTGAGCTGAAGGCTCTGAGCTCAGAG	1499
Db	2208	AGAGTGAAGGCTGTGCTTATGTCAGAGAGGTGAGCTGAAGGCTCTGAGCTCAGAG	2267
Qy	1500	ATCCCTCTGAGAGAGGACCTTGTCTTCAATCAAGATGAGATGAGATGAGATGAGATG	1559
Db	2268	ATCCCTCTGAGAGAGGACCTTGTCTTCAATCAAGATGAGATGAGATGAGATGAGATG	2327
Qy	1560	TTTTGAGCAAGCAAGATGAGGATGTCACAGAGATGACAAAGCACTGAGCTTCTCA	1619
Db	2328	TTTTGAGCAAGCAAGATGAGGATGTCACAGAGATGACAAAGCACTGAGCTTCTCA	2387
Qy	1620	TGATATCAGAGAGAGAGGCGGAGGCTTCAAGAAATCAAGAGCAGAGATGACAGGCTCA	1679
Db	2388	TGATATCAGAGAGAGAGGCGGAGGCTTCAAGAAATCAAGAGCAGAGATGACAGGCTCA	2447
Qy	1680	AGTGAAGAAATGAGTGAATGATGAATCAAGTTGAAGAGATCTTGTCTCAGCAAGAG	1739
Db	2448	AGTGAAGAAATGAGTGAATGATGAATCAAGTTGAAGAGATCTTGTCTCAGCAAGAG	2507
Qy	1740	ACGAGATGATCTTCAAGAACTGATGATGATGATGATGATGATGATGATGATGATG	1799
Db	2508	ACGAGATGATCTTCAAGAACTGATGATGATGATGATGATGATGATGATGATGATG	2567
Qy	1800	GCGGAAAGCAAGAAATGTCAGATTAATCTGTTGAAGGCTTAAGATCAAGGAGGCTGA	1859
Db	2568	GCGGAAAGCAAGAAATGTCAGATTAATCTGTTGAAGGCTTAAGATCAAGGAGGCTGA	2627
Qy	1860	AGTGGAGAAATATGCGAACTGAGAGAAATCAATGCTGAGCAGCAGACTCAAAATTCAGA	1919

Db 2628 AGTGGAGAAATATTCCTGAACTGGAGAAAGTCAATGCTTGAGCAGCACTGAAAGTCCAGGA 2687  
Qy 1920 GCTCCAAAGAACTGGAGAAAGCTGTAAAGCCAGACAGAGGCCACCGAGCTGTGCA 1979  
Db 2688 GCTCCAGAGAAAGCTGGAAGAGGCTGTAAAGCCAGACAGAGGCCACCGAGCTGTGCA 2747  
Qy 1980 GAATATCCGCCAGGCAAGAGAGCGAGCCAGAGGGAGCTGTGAGAAAGCTGCAGAACCGAGA 2039  
Db 2748 GAATATCCGCCAGGCAAGAGAGCGAGCCAGAGGGAGCTGTGAGAAAGCTGCAGAACCGAGA 2807  
Qy 2040 GGAATCTTGTGAAGCATGAGAAAGAGCTGTGAGAGAGCTGAGAGAAAGCGCCGCTATTTCT 2099  
Db 2808 AGACTCTCCAGAGGCGATTAAGAGAGCTGTGAGAGCGAGAGAAAGCGCCGCTATTTCT 2867  
Qy 2100 GGAGAACAAAGTAAAGAGCTGAGAGCACTGAGAGCTGAGAGAAACAGACTGAGAGATGA 2159  
Db 2868 GGAGAACAAAGTAAAGAGCTGAGAGCACTGAGAGCTGAGAGAAACAGACTGAGAGATGA 2927  
Qy 2160 CATCCAGCAAAAATCCCAACAGATCCAGAGAGTGGCTGATTAATTTCTGAGCTCGAAGA 2219  
Db 2928 CATCCAGCAAAAATCCCAACAGATCCAGAGAGTGGCTGATTAATTTCTGAGCTCGAAGA 2987  
Qy 2220 GAAACATCGGAGAGGCCCAAGTCTCAGCCCAAGCACTGAGAGTGCACCTGAGAACAGAGAGA 2279  
Db 2988 GAAACATCGGAGAGGCCCTCAGGCTCTCAGCTCAACATTAAGAGTGCACCTGAGAACAGAGAGA 3047  
Qy 2280 GCAGACATATGAGAGAAAGATTTAAAGTGTGACATAGATTAAGAAAGACCTGGCTGA 2339  
Db 3048 ACAGACATACGAGAGAAAGATTCMAAGTATTGGACATCAGATTAAGAAAGAGACCTGGCCGA 3107  
Qy 2340 CAAGGAGACACTGGAGAACATGATGACAGAGACAGAGAGAGAGGCCCATGAGAGAGGCA 2399  
Db 3108 CAAGGAGAGCTTGGAGAACATGATGACAGAGACAGAGAGAGAGGCCCATGAGAGAGGCA 3167  
Qy 2400 AATTCTCAGCGAACAGAGAGCGATGATCATGCTATGATTCAGATTCAGATTCAGATTCCTTGA 2459  
Db 3168 GATCCTCAGCGAGAGAGAGGAGATGATCAAGCGATGATTCAGATTCAGATTCCTTGA 3227  
Qy 2460 ACAGAGATTTGTGAACCTGTCTGAAGCCAAATTAATTTGCAAGTCTTTTAC 2519  
Db 3228 GCAGAGATCGTGGAGTGTGCGAAGCCAAACAGCTTCGGGCAACAGACGTCTTTCAC 3287  
Qy 2520 CCAAGAGAACATGAAGGCCCAAGAGAGATGATTTCTGAACCTGAGGCAAGAAATTTTA 2579  
Db 3288 CCAAGAGAACATGAAGGCCCAAGAGAGATGATTTCTGAAGCTGAGGCAAGAAATTTTA 3347  
Qy 2580 CTTGAGACACAGGCTGGAGAAATTTGAGAGCCAGAACCGAAACTGAGAGAGCTTGA 2639  
Db 3348 CTTGAGAGACGAGGCCGAGAAAGCTGAGAGCCAGAACCGAAAGCTGAGAGAGCTTGA 3407  
Qy 2640 GAAATACGCCAACCAAGACCAAGATGATTTCTGAACCTGAGGCAAGAAATTT 2699  
Db 3408 GAAATACGCCAACCAAGATGATTTCTGAAGCTGAGGCAAGAAATTT 3467  
Qy 2700 GCGGAGAGTCACTGAGAGCAAGAGAGCAAGAACTGAGGCTCAAGCGCCGCTCAACGA 2759  
Db 3468 GAGGAGAGTCACTGAGAGCAAGAGAGCAAGAACTGAGGCTCAAGCGCCGCTCAACGA 3527  
Qy 2760 GCTACAGCTCTCCCTGAGAGAGCGAGTCAAGTTGAACAGCTTGAAGCTTGAAGCTTGAAGCGGC 2819  
Db 3528 GCTACAGCTCTCCCTGAGAGAGCGAGTCAAGTTGAACAGCTTGAAGCGGC 3587  
Qy 2820 GGCCCTGAGAGAGCGAGCTTGGCAGCGAGGAGAAACAGAGCTGAGAGAGACAGAGAGAGC 2879  
Db 3588 AGCCCTGAGAGAGCGAGCTTGGCAGCGAGGAGAAACAGAGCTGAGAGAGACAGAGAGAGC 3647  
Qy 2880 TGAAGAGAGATCCAGAGCACTCAGGCAATAGAGATGAATCCAGGCGAAATTTGATGC 2939  
Db 3648 GGAAGAGAGATCCAGGCGCTCAGGCAATCCAGGAGATGAATCCAGGCGAAATTTGATGC 3707  
Qy 2940 TCTTCTGAACAGCTGTCTGTATCAAGACTTGAAGAGAGAGCTTAACAGCTGAGAGC 2999

Db 3708 CCTTGCAACAGCTGACCGGTCAATCACAGCTTGAGAGAGAGCTGAACCAAGCTCACCGA 3767  
Qy 3000 GGAACAAGCTGAACCTCAACCAACAACTTCACTTGTCCAAACCAACTGATGAGGCTTC 3059  
Db 3768 GGAACAAGCTGAGCTCAACCAACAACTTCACTTGTCCAAACCAACTGATGAGGCTTC 3827  
Qy 3060 TGGCGCCAAAGAGATTTGTACAACTGCGAAGTGAAGTGAACCAATCTCCGCGGAGAT 3119  
Db 3828 CGGGGCGAATGACGAGATTTGTACAGCTGCGAGATGAAGTGAACCAATCTCCGCGTAAAT 3887  
Qy 3120 CACGGAACGAGATGACAGCTTACAGCCAGAGAACAGATGAGGCTCTGAGAGCAAC 3179  
Db 3888 CACGGAACGAGATGACAGCTTACAGCCAGAGAACAGATGAGGCTCTGAGAGCAAC 3947  
Qy 3180 GTGCAACATGCTGAGAGAGAAACAGTCAATGATTTGAGAGGCTTAAACATGAGCTGTAGA 3239  
Db 3948 ATGCAACATGTTGAGAGAGCAAGTCTTGAACCTGAGAGGCTTAAACATGAGCTGTAGA 4007  
Qy 3240 AAAAGAGCGGCGAGTGGAGAGGCTTGGAGAGAGCTCTGAGGTGATGAGAAATCCAGTTGA 3299  
Db 4008 GAAAGAGCGGCGAGTGGAGAGGCTTGGAGAGAGCTCTTGGCGAGAGAGTCCAGTTGA 4067  
Qy 3300 GTTGGGTTGAGAGAGCTGAGAGAGATGCTGGAACCCAGAGAAACAGAGAGGCGAGAGC 3359  
Db 4068 GTTGGGTTGAGAGAGCTGAGAGAGATGCTGGAACCCAGAGAAACAGAGAGGCGAGAGC 4127  
Qy 3360 CGATCAGCGAGATCAACGAGTCTCGCAGAGTGGAGAGCTGGAGAGAGAGAGAGAGC 3419  
Db 4128 CGATCAGCGAGATCAACGAGTCTCGCAGAGTGGAGAGAGTGGAGAGAGAGAGAGAGC 4187  
Qy 3420 TGAATTTCTGCTTGCAGAGAGCTCTGAAAGAGAGAGAGTGAAGGCGGAGAGCTCTC 3479  
Db 4188 CGAGATTTCTGCTTGCAGAGAGCTCTGAAAGAGAGAGAGTGAAGGCGGAGAGCTCTC 4247  
Qy 3480 TGAAGCTCAATGACCTGAGAGAGAGAGAGAGTGTGCTTGAATGAATGCCGAAGCTT 3539  
Db 4248 GGAAGCTCAATGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4307  
Qy 3540 ACAGAGAGAGTGGAGACTGAGACGAGAGCTCAACAGAGAGCTTGGAGAGAGAGAGAGC 3599  
Db 4308 ACAGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4367  
Qy 3600 ATTACAGAGAGAGATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 3659  
Db 4368 ATTACAGAGAGAGATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4427  
Qy 3660 AGAAGCTTGAATGGGCTGATCTACTGAAGACAGAAAGAGAGAGAGAGAGAGAGC 3719  
Db 4428 GAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4487  
Qy 3720 GGAAGAGAGTTCATGATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 3779  
Db 4488 GGAAGAGAGTTCATGATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4547  
Qy 3780 ACAAACCAACTCATGATTTTCTGCAAGCCAAATGAGAGAGAGAGAGAGAGAGC 3839  
Db 4548 GCAAACCAACTCATGATTTTCTGCAAGCCAAATGAGAGAGAGAGAGAGAGAGC 4607  
Qy 3840 GGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 3899  
Db 4608 GGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4667  
Qy 3900 AGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 3959  
Db 4668 GAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4727  
Qy 3960 TGCCCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4019  
Db 4728 TGCCCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4787  
Qy 4020 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4079  
Db 4788 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 4847









Dp	961	CCGA-----	964
QY	2138	GAAGAAACGACTGAGAGTGAATCTCCAGACAAAATCCCAACAGATCCAGAGATGCTG	2197
Dp	965	-----	964
QY	2198	ATTAATTTCTGGAGCTCGAGAGGAAACATCCGGAGGCCAAGTCTCAAGCCAGACCTAG	2257
Dp	965	-----GAGCTGGAGAGAAACATCCGGAGGCTCAAGTCTCAAGCTCAACATCTAG	1014
QY	2258	AAGTCACTTGAAACAGAAAGGAAAGCAGCACTATGAGAAAAGATTAACTGTGGACATC	2317
Dp	1015	AAGTCACTTGAAACAGAAAGGAAAGCAGCACTATGAGAAAAGATTAACTGTGGACATC	1074
QY	2318	AGATTAAGAAAGACCTGGCTGACAGAGAGACACTGAGAAAATGATGTAGAGACACAGAG	2377
Dp	1075	AGATTAAGAAAGACCTGGCTGACAGAGAGACACTGAGAAAATGATGTAGAGACACAGAG	1134
QY	2378	AGAGAGCCCATAGAGAAAGGCAAAAATTTCAAGCGAACAGAAAGCGATGATCAATCTATAG	2437
Dp	1135	AGAGAGCCCATAGAGAAAGGCAAAAATTTCAAGCGAACAGAAAGCGATGATCAATCTATAG	1194
QY	2438	ATTCCAAATCAGATTCCTCTGGAACAGAGGATTTGTGAACTGTCTGAGCCATTAATTG	2497
Dp	1195	ATTCCAAATCAGATTCCTCTGGAACAGAGGATTTGTGAACTGTCTGAGCCATTAATTG	1254
QY	2498	CAGCAAAATAGCAGTCTTTTATCCCAAGAAACATGAAGGCCACAGAAAGATGATTTCTG	2557
Dp	1255	CGGCAAAACAGCAGTCTTTTATCCCAAGAAACATGAAGGCCACAGAAAGATGATTTCTG	1314
QY	2558	AACTCAGGCAACAGAAATTTTATCTGTGAGACACAGCTGGAGAGTTGGAGGCCAGAAC	2617
Dp	1315	AACTCAGGCAACAGAAATTTTATCTGTGAGACACAGCTGGAGAGTTGGAGGCCAGAAC	1374
QY	2618	GAAAACCTGAGAGAGAGCTGAGAGAAAGATCAAGCCACAGACACAGATGACAAAGATCGGC	2677
Dp	1375	GAAAACCTGAGAGAGAGCTGAGAGAAAGATCAAGCCACAGACACAGATGACAAAGATCGGC	1434
QY	2678	TGCTGGAACTGAGAGCAAAATTTGCGCGAGATCACTGATCTAGACACAGAGAGCAGAACTGG	2737
Dp	1435	TGCTGGAACTGAGAGCAAAATTTGCGCGAGATCACTGATCTGATGACACAGAGAGCAGAACTGG	1494
QY	2738	AGCTCAACCGCAGCTTCACAGAGCTACAGACTCTCCCTGACAGAGCGGAGTCAACATTTGA	2797
Dp	1495	AGCTCAACCGCAGCTTCACAGAGCTACAGACTCTCCCTGACAGAGCGGAGTCAACATTTGA	1554
QY	2798	CAGCCCTCAGAGCTGCAAGGCGGCGCTGTGAGAGCAGACTTCGCGCAGCGAAGACAGAGC	2857
Dp	1555	CAGCCCTCAGAGCTGCAAGGCGGCGCTGTGAGAGCAGACTTCGCGCAGCGAAGACAGAGC	1614
QY	2858	TGGAAGAGCCACAGCAGAGAGTGTGAAGAGAGATCAGAGCACTCACCGCACTAAGAGATG	2917
Dp	1615	TGGAAGAGCCACAGCAGAGAGTGTGAAGAGAGATCAGAGCACTCACCGCACTAAGAGATG	1674
QY	2918	AAATCCAGCAGCAAAATTTGATGCTCTTCGTAACAGCTGACTGATATCAACAACCTGAGAG	2977
Dp	1675	AAATCCAGCAGCAAAATTTGATGCTCTTCGTAACAGCTGACTGATATCAACAACCTGAGAG	1734
QY	2978	AGCAGCTTAACCAAGCTGACGAGAGCAACAGCTGAACTCAACCAACCAAACTTCACTTGT	3037
Dp	1735	AGCAGCTGTAACCAAGCTGACGAGAGCAACAGCTGAACTCAACCAACCAAACTTCACTTGT	1794
QY	3038	CCAAACCACTGATGAGGCTCTTGCGCGCAACGACGAGATTGTACACTGGAAGTGAAG	3097
Dp	1795	CCAAACCACTGATGAGGCTCTTGCGCGCAACGACGAGATTGTACACTGGAAGTGAAG	1854
QY	3098	TGAGCCATCTCCGCGGAGATTCACGGAACAGAGATGCAAGCTTACAGCCAGAGAGAAA	3157
Dp	1855	TGAGCCATCTCCGCGGAGATTCACGGAACAGAGATGCAAGCTTACAGCCAGAGAGAAA	1914
QY	3158	CGATGAGAGCTCTTGAAGACCAAGTGCACATGCTGAGAGAAACAGCTCATGATTTGAGAG	3217
Dp	1915	CGATGAGAGCTCTTGAAGACCAAGTGCACATGCTGAGAGAAACAGCTCATGATTTGAGAG	1974

QY	3218	CCCTAAACGATGACCTGCTAGAAAABAAGCGCGCATGTGGAGAGCCTCGAAGACGCTCCTCG	3277
Db	1975	CCCTGAAGAGACACTGCTGAGAAAGAGCGCCAGTg3gA93CCTGCGAGGCGCTCCTTG	2034
QY	3278	GTGATGAGAAATCCACAGTTTGATGATGCTGGGCTTCAGAGCTGCAGAGATGCTGGACACCG	3337
Db	2035	GGAGCGAAGATCCCAAGTTCCAGTTCGAGGTTCCAGAGCTACAGAGATGCTGGACACCG	2094
QY	3338	AGAAAACAGACGAGGCGCAGAGCCGATCAGCGGATTCACCGAGTCTCGCCAGGTGGAGC	3397
Db	2095	AGAAAGCAGAGAGGCGCTAGGCGCCATCAGCGGATTCACCGAGTCTCGCCAGGTGGAGT	2154
QY	3398	TGGCAGTGAAGAGAGACAAAGGCTGAGAGTTCGCTCTGCAGACAGAGCTTCAAAGACAGA	3455
Db	2155	TGGCGGTGAAGAAACAAGGCGAGATTTCTTGCTCTGCACAGGCTCTCAAAGACAGA	2214
QY	3458	AGCTGAAGCGGAGAGCCTCTCTGACAAAGCTCAATGACTTGGAGAAAGACATGCTATGC	3517
Db	2215	AGCTCAAAAGCCGAGAGCCTGTCCGACAAGCTCAACGACTTGGAGAAAGACAGCGCATGC	2274
QY	3518	TTGAAATGAATAGCCCGGAAAGCTTACAGAGAGCTGGAGACTGAAACGAGAGCTCAACAGA	3577
Db	2275	TGGAGATGAAGCGCCCGAGCTTACAGAGAACTAGAGACAGAGCGGAGCTCAACAGA	2334
QY	3578	GGCTTCTGGAAGACCAAGCCAAATTAACAGACAGATGAGCTTCAGAAAAATCACATTT	3637
Db	2335	GGCTTCTGGAAGACCAAGCCAAATTAACAGAGACAGATGAGCTTCAGAAAGACACATCT	2394
QY	3638	TCCTGTTGATCTCAGAGACTGCAGAGAGCTTGATTCGGGCTGATCTACTGAAACAGAAA	3697
Db	2395	TCACACTGACCCCAAGGCGCTGCAGAGGCGCTGGACCGGCGGATCTGCTGAAACAGAAA	2454
QY	3698	GAAGTGACTTGAAGATACGCTGGAAAAACATTGAGTTCTCTATTCATGAAAAAGTGA	3757
Db	2455	GGAGCGACTGGAATACAGCTGAAAAACATTGAGTTCTCTACTCTCAGAGAAAGTGA	2514
QY	3758	AAATGGAAGGCACTATTTCTTCAACAAAACAAACTCATTTATTTTCTGCAAGCCAAAATGG	3817
Db	2515	AAATGGAAGGCACTATCTCTCAGCAAAACAAACTCATTTATTTCTGCAAGCCAAAATGG	2574
QY	3818	ACCAAGCTGCTAAAAAAGAAAAAGTTCCTCTGCAGTACATAGAGCTGAAGCTGGCCCTGG	3877
Db	2575	ACCAAGCTGCTAAAAAAGAAAAAGTTCCTCTGCAGTACATAGAGCTGAAGCTGAAGCTGG	2634
QY	3878	AGAAAGGAAAGAGCTGCTGCTGTGCAGAGCTAGAGAGAGCCCTTCAGAAAGACCCGACTGAGC	3937
Db	2635	AAAAGGAAAGAGCCCGATGGCGCGGAGCTTGAGAGAGGCTTTCAGAAAGACCCGACTGAGC	2694
QY	3938	TCGCGTCCGCCCGGAGAGAGCTGCGCACCGCAAAAGCAACGACACCCATCCACACCCATCCA	3997
Db	2695	TCGCGTCCGCCCGGAGAGAGCTGCGCACCGCAAAAGCAACGACACCCGACTGCTA	2754
QY	3998	CGCCAGGCAACCGCAGAGGAGAGATGCGCAATGCTCGGCAATCGGCGGCTGCGCAGAGAGAC	4057
Db	2755	CGCCAGGCACTGCGCAGAGGAGAGATGCGCAATGCTGCGCAATTTGGCGGCTGCGCAGAGAC	2814
QY	4058	AGCCCAAGTCCATGAGCCTGTGCGCCCGCCATCCAGCCGCGAGAAAGAGTCTTCAATC	4117
Db	2815	AGCCCAAGTCCATGAGCCTGTGCGCCCAATCCATCCAGCCGCGAGAAAGAGTCCGCAATC	2874
QY	4118	CAGAGGAATTTAAGTCCGCGCTTAAAGAAAGCATGACCCACAATATTTCTCACCGATTCA	4177
Db	2875	CAGAGGAATTTAAGCGCGCGTCTGAAAGAGGCAATGACCAACAATCCCTCACCGATTTA	2934
QY	4178	ACGTGAGACTGAACATGCGAGCCCAAAAGTGTGCTGTGCTTGATTAACGTGCACTTTG	4237
Db	2935	ATGTGGGCGTGAACATGAGAGCCACCAAGTGGCGCGTGTCTGGAATACGTGCACTTTG	2994
QY	4238	GAGCCGAGCATCCAAATGTCTGGAATGTCAAGTATGTGTCACTCCCAAGTGTCCACGT	4297
Db	2995	GAGCCGAGCATCCAAATGTCTGGAATGTCAAGTATGTGTCACTCCCAATGTCTCACT	3054

QY 4298 GCTTGCAGACACCTGGCTTGCCTGTAATATSCACACACTTCAACGAGGCTTCT 4357  
 DB 3055 GCTTGCAGACACCTGGCTTGCCTGTAATATSCACACACTTCAACGAGGCTTCT 3114  
 QY 4358 GCCGTGACAAATGAACTCCCAAGTCTCCAGACCAAGAGCCACAGACCTTGCACC 4417  
 DB 3115 GCCGTGACAAATGAACTCCCAAGTCTCCAGACCAAGAGCCACAGACCTTGCACC 3174  
 QY 4418 TGGAAGGTGATGAAGTGGCCAGGATTAACAAGAGACAGACAGCTGGGACAG 4477  
 DB 3175 TGGAAGGTGATGAAGTGGCCAGGATTAACAAGAGAGAGAGAGAGCTGGGACAG 3234  
 QY 4478 AGTACATTTGCTGAGAGGATCAAAAGTCTCATTTATGACATGAGACAGAGAGCTG 4537  
 DB 3235 AGTACATTTGCTGAGAGGATCAAAAGTCTCATTTATGACATGAGACAGAGAGCTG 3294  
 QY 4538 GACAGAGGCGGTGAGAGATTTGAGCTGTGCTTCCAGCGGAGTGTATTTATG 4597  
 DB 3295 GACAGAGGCGGTGAGAGATTTGAGCTGTGCTTCCAGCGGAGTGTATTTATG 3354  
 QY 4598 GTGCGCTTGTGCTTCCAGCTGCAATACAGCCAAAGACAGTGTCCATATCA 4657  
 DB 3355 GTGCGCTTGTGCTTCCAGCTGCAATACAGCCAAAGACAGTGTCCATATCA 3414  
 QY 4658 AGATGAAATCTCACCCGACACCACTGCTGCGGAGAAACCTTCACTTCTAGCTC 4717  
 DB 3415 AGATGAAATCTCACCCGACACCACTGCTGCGGAGAAACCTTCACTTCTAGCTC 3474  
 QY 4718 CCACTTCCCTGACCAACAGCGGTGAGTACCGCTTGAATCAAGTGTGAGAGTGG 4777  
 DB 3475 CCACTTCCCTGACCAACAGCGGTGAGTACCGCTTGAATCAAGTGTGAGAGTGG 3534  
 QY 4778 GAGTTTCTAGAGGAAAGAGAGAGCTGATGCTAACTGCTTGGAAACCTCTGTAAC 4837  
 DB 3535 GAGTTTCTAGAGGAAAGAGAGAGCTGATGCTAACTGCTTGGAAACCTCTGTAAC 3594  
 QY 4838 TGGAAGGTGATGAAGTGGCCAGGATTAACAAGAGAGAGAGAGAGTGGT 4897  
 DB 3595 TGGAAGGTGATGAAGTGGCCAGGATTAACAAGAGAGAGAGAGTGGT 3654  
 QY 4898 TGGTGGGACACGAGAGAGGCTCTACGCTGATGCTTGAAGAAACCTCTAACCCATG 4957  
 DB 3655 TGGTGGGACACGAGAGAGGCTCTACGCTGATGCTTGAAGAAACCTCTAACCCATG 3714  
 QY 4958 TCCAGAGATTTGAGACAGCTTTCGAATTTATTTATCAAGAGCTGGAGAGCTACTCA 5017  
 DB 3715 TCCAGAGATTTGAGACAGCTTTCGAATTTATTTATCAAGAGCTGGAGAGCTACTCA 3774  
 QY 5018 TGATAGCAGAGAGAGAGAGGAGCTGTGTCTTGTGAGAGTGAAGAGAGAGTGGC 5077  
 DB 3775 TGATAGCAGAGAGAGAGAGGAGCTGTGTCTTGTGAGAGTGAAGAGAGTGGC 3834  
 QY 5078 TGACCCAGTCCACCTGCTGCTGCCAGCCGACATCTCACCAACATTTTGAAGCTGTCA 5137  
 DB 3835 TGACCCAGTCCACCTGCTGCTGCCAGCCGACATCTCACCAACATTTTGAAGCTGTCA 3894  
 QY 5138 AGGGCTGCACTGTTTGGGGAGAGCAAGTTGAGAGAGGGGCTGATCTGTGAGAGCA 5197  
 DB 3895 AGGGCTGCACTGTTTGGGGAGAGCAAGTTGAGAGAGGGGCTGATCTGTGAGAGCA 3954  
 QY 5198 TGCCCAAGCAAGTGTCTTCTCCGCTACACAGAAACCTCAGCAATATCTGATCCGA 5257  
 DB 3955 TGCCCAAGCAAGTGTCTTCTCCGCTACACAGAAACCTCAGCAATATCTGATCCGA 4014  
 QY 5258 AAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5317  
 DB 4015 AAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4074  
 QY 5318 TTGGAACCAATATTTCTAGAAATGACATGAGAGAGAGAGAGAGAGAGAGAGAGAG 5377  
 DB 4075 TTGGAACCAATATTTCTAGAAATGACATGAGAGAGAGAGAGAGAGAGAGAGAGAG 4134  
 QY 5378 ATTAGAATGACATTTCTTGGGACCTGTGTGTGTTGGCCCTCTTCAAGAGCTTCCCTG 5437

DB 4135 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4194  
 QY 5438 TCTCAATGCTGAGAGTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5497  
 DB 4195 TCTCAATGCTGAGAGTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4254  
 QY 5498 AATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5557  
 DB 4255 AATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4314  
 QY 5558 GTGCTTACCTTTGAGCTTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5617  
 DB 4315 GTGCTTACCTTTGAGCTTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4374  
 QY 5618 CACTGGAAGTAAATGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5677  
 DB 4375 CACTGGAAGTAAATGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4434  
 QY 5678 TGGAATCCCAACCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5737  
 DB 4435 TGGAATCCCAACCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4494  
 QY 5738 GTCCTCAATACAGAGTAAATTAAGAGTCAATTTGCTGCAAGAGAGAGAGAGAGAG 5797  
 DB 4495 GTCCTCAATACAGAGTAAATTAAGAGTCAATTTGCTGCAAGAGAGAGAGAGAGAG 4554  
 QY 5798 CCGGACCTGAAACACACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5857  
 DB 4555 CCGGACCTGAAACACACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4614  
 QY 5858 CACCAAGCTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5917  
 DB 4615 CACCAAGCTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4674  
 QY 5918 GCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5971  
 DB 4675 GCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4734  
 QY 5972 CCGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6031  
 DB 4735 CCGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4794  
 QY 6032 GGAATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6091  
 DB 4795 GGAATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4854  
 QY 6092 GCGGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6151  
 DB 4855 GCGGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4914  
 QY 6152 AGTCTTCAATATAA 6165  
 DB 4915 AGTCTTCAATATAA 4928

RESULT 15  
 MNU3904  
 LOCUS DEFINITION 5019 bp mRNA linear ROD 30-DEC-1995  
 ACCESSION U39904  
 VERSION U39904.1 GI:107973  
 KEYWORDS signal transduction; rho; coiled-coil; leucine-zipper; pleckstrin homology; zinc finger.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCES  
 AUTHORS Madanile, P., Puriyashiki, T., Reid, T., Ishikaki, T., Wetanabe, G.,  
 Motil, N., and Narumiyu, S.  
 A novel partner for the GTP-bound forms of rho and rac  
 FEBS Lett. 377 (2), 243-248 (1995)

MEDLINE 96128238  
 PUBMED 8543060  
 REFERENCE 2 (bases 1 to 5019)  
 AUTHORS Madaule, P., Furuyashiki, T., Reid, T., Ishizaki, T., Watanabe, G.,  
 Motii, N. and Natumiya, S.

TITLE Direct Submission  
 JOURNAL Submitted (02-NOV-1995) Pascal Madaule, Department of Pharmacology,  
 Faculty of Medicine, Kyoto University, Yoshida, Konoe, Sakyo-ku,  
 Kyoto, 606, Japan

FEATURES  
 source Location/Qualifiers  
 1..5019

/organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /tissue\_type="brain"  
 /dev\_stage="adult"

misc\_feature  
 /note="Multiple alternative splices affect the N-terminal region. The 5' end may be replaced by another sequence, resulting in a protein with an N-terminal extension"  
 182..4975

CDS  
 /note="binds to the GTP-bound forms of p21 rho and rac, according to two hybrid system and overlay assay; putative rho/rac effector"  
 /codon\_start=1  
 /product="citrin"  
 /protein\_id="AAC52341.1"  
 /db\_xref="GI:1079734"

/translation="MLGEEAMMEOEENTLRHRRVEAVLSQKEVELKASRTSIL  
 /EODLTYTESCSLSKSLBOARMSVSOEDKALQDLIDIEGRKLOELKEDEYQOV  
 EEMIMNMOLEEDVARSRRSDLYESLEERSRLAEPRKKAECOHKIMKMDOKGP  
 EYEGSKLEKINAEOCKLOELKELKAKASTAETELLQNRKAKARELLEKTH  
 NREDSBGTIKKLVLEERBRHSLENTKYLLETMRERBNLKDIDITKSQIQOMAKI  
 LLEEKREAOVSAQHLVHLKQKEQHEKIKVLDNQIKDLADESEMMKQEMREI  
 BAEHKGKILEOKAMINAMDSKISLEORVELSEANKLAANSLETOENKAOEEMI  
 SEIRQKFLYETQAGKLEQNRKLEBOLKISHQSDSKSLLETSLREAVLEIHR  
 QKLEKROLTELOESLOERESQALQAAALAESOLROKXTELETETRAVEEILAL  
 TAHRDEIQRKPDALRNSCTVITDLERQNTEDNLEINNONPYLSKQDASGAND  
 IVQIRSEVDLREIREMOLTSQKQMTALTTCTMEBEQULDEBALDELLEKER  
 QWEARSVLDGERSQFECRVELQRMIDTEKQSRABDQITSRQVLEAVEHAE  
 ILAQALKEQKRLKASLSKDLKNDLEKAMLEMMARSLQKLETREKORLEBOA  
 KLOQMDLQKRLIFRLTQSGEALDRADLKTERTSLEYQLENIQVLYSHEKVMQ  
 ISOOTKLIDFLQAMDQAPAKKVPLOYNLSKLAEKARCALEBALOKTEIERS  
 ABEAAHRAKTDHPSTPATPAQOOLMSAIVSRPHOPAMSLAPSSRRSESTP  
 BEPSRLKEMRHNIPIRFRVGMNMTAKVCLDTPHROKSKLECOQYMKPECS  
 TCEPATCGLAETATHTTEAFCDKAKMSPELQSKPESSSLHLEGMKRENRGQOG  
 WDRKTVLESGKVIYDNEAREAGORVESEFCLPQGVDSHGAGASBLANTAD  
 VPIYLKESHPHTTQWPGRTLYLLAPFPBKQWVYALNLSLTHIPGIGAVQIY  
 LQNSLLKLEBDRLDNMCPLPESDQVVLCTEBGLVALNLSLTHIPGIGAVQIY  
 IIKDLKILMIAGERALCLVDYKVKOSLAQSLPAQVPSNIFEAVGCHLPAAG  
 KIENSLCTICAMPKSVILRYNDLSKYCIKREIETSEPCSHIFPNYSILITKNEY  
 EIMKQYTLDEPLDKNDHSLAPAVFASSSSPVSIQVANSAGOREYILLCFEBCV  
 VDSYGRSRTDCLKSRPLAPAYRBEPLFYTHPNSLEVIETQARSLSAPAYLEI  
 PNEYRGLPALSGLIYASLYQDLRTICKNVLKESGEORVSTERSKNGKGP  
 PLYNEHTKRVASSPAPPEGSHREBSTRHRYRDEGRTELRLRDSPLRPERESP  
 GRMLSTRERSVGLFEDSSRGLPAAVATPLSQVWKVDDSSV"  
 887..1012  
 /note="this exon is dispensable"

BASE COUNT 1399 a 1350 c 1466 g 804 t  
 ORIGIN

Query Match 65.2%; Score 4018.2; DB 10; Length 5019;  
 Best local Similarity 90.3%; Pred. No. 0;  
 Matches 4309; Conservative 0; Mismatches 458; Indels 6; Gaps 1;

QY 1399 AAGATGAGCAGCAATGACCCGGTTACATCGAGAGATGTCACAGGTGGAGGCTTGCTT 1458  
 Db 203 ATGATGAGCAGCAATGACCCGGTTACATCGAGAGATGTCACAGGTGGAGGCTTGCTT 262  
 QY 1459 AGTCAGAGAGGTGGAGCTGGAAGGCTTGAAGATCTGAGATCTCCTCTGAGCAGAGC 1518  
 Db 263 AGTCAGAGAGGTGGAGCTGGAAGGCTTGAAGATCTGAGATCTCCTCTGAGCAGAGC 322

QY 1519 CTTGCTACCTACATCAGAAATGAGTAGTCTTAAAGCAAGTTTGGACAGACGATG 1578  
 Db 323 CTTGCTACCTACATCAGAAATGAGTAGTCTTAAAGCAAGTTTGGACAGACGATG 382  
 QY 1579 GAGGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCCATGATATAGAGACAGAC 1638  
 Db 383 GAGGTGTCCAGAGAGATGACAAAGCACTGACGCTTCTCCATGATATAGAGACAGAC 442  
 QY 1639 CGGAAGCTCCAGAAATCAAAAGAGAGAGTAGACAGGCTCAAGTGAAGAAATGAGTTG 1698  
 Db 443 CGGAAGCTCCAGAAATCAAAAGAGAGAGTAGACAGGCTCAAGTGAAGAAATGAGTTG 502  
 QY 1699 ATGATGAATCAAGTTGAAAGAGATCTTGTCTCAGCAGAAAGACGAGTATCTTACGAA 1758  
 Db 503 ATGATGAATCAAGTTGAAAGAGATCTTGTGTCAAGCCGACAGACGAGATCTTACGAG 562  
 QY 1759 TCTGAGCTGAGAGATCTTCGGCTTCTGCTCTGAAGATTCAGCGGAAAGCCACAAATGT 1818  
 Db 563 TCTGAGCTGAGAGATCTTCGGCTTCTGCTGCGCCGAGGAATTCAGCGGAAAGCCAAATGT 622  
 QY 1819 CAGCATTAACCTGTTGAAGGCTTAAGGATCAAGGAAAGCCTGGAAGTGGAGAAATATGCAA 1878  
 Db 623 CAGCATTAACCTGTTGAAGGCTTAAGGATCAAGGAAAGCCTGGAAGTGGAGAAATATGCAA 682  
 QY 1879 CTGAGAAAGATTAATGCTGAGCAGCAGCTCAAAATTCAGAGCTTCAAGAGAAATCTGAG 1938  
 Db 683 CTGAGAAAGATTAATGCTGAGCAGCAGCTCAAAATTCAGAGCTTCAAGAGAAATCTGAG 742  
 QY 1939 AAGGCTGTAAAGCCAGACAGGAGCCACCGAGCTGCTGCAAAATATCCGCGACGCAAG 1998  
 Db 743 AAGGCTGTAAAGCCAGACAGGAGCCACCGAGCTGCTGCAAAATATCCGCGACGCAAG 802  
 QY 1999 GAGCAGCCGAGAGGAGCTGAGAGAGCTGCAAGCCAGAGATTTCTTGAAGCAATC 2058  
 Db 803 GAGCAGCAGAGCGGAGAGCTGAGAGAGCTGCAAGCCAGAGATTTCTTGAAGCAATC 862  
 QY 2059 AATAAGAGCTGTGTGAAGCTGAGAGAGCGCCGCTTCTGGAAGCAAGATTAAGGA 2118  
 Db 863 AATAAGAGCTGTGTGAAGCTGAGAGAGCGCCGCTTCTGGAAGCAAGATTAAGGA 922  
 QY 2119 CTGAGAGCAGTGAAGCTGAGAGAAACAGACTGAAGATGACATCCAGACAAATCCCA 2178  
 Db 923 CTGAGAGCAGTGAAGCTGAGAGAAACAGACTGAAGATGACATCCAGACAAATCCCA 982  
 QY 2179 CAGATCCAGAGATGCTGATTAATTTCTGAGACTGCAAGAGAAACATCCGAGAGCCCA 2238  
 Db 983 CAGATCCAGAGATGCTGATTAATTTCTGAGACTGCAAGAGAAACATCCGAGAGCCCA 1042  
 QY 2239 GTCTAGGCCAGACCTTGAAGTGCACCTGAAACAGAAAGACAGACCTATAGAGAAAG 2298  
 Db 1043 GTCTAGGCCAGACCTTGAAGTGCACCTGAAACAGAAAGACAGACCTATAGAGAAAG 1102  
 QY 2299 ATTAAAGTGTGGAATCAATCAATTAAGAAAGACCTGCTGACAGAGACACTGAGAAC 2358  
 Db 1103 ATTAAAGTGTGGAATCAATCAATTAAGAAAGACCTGCTGACAGAGACACTGAGAAC 1162  
 QY 2359 ATGATGAGAGACAGAGAGAGAGGCGCATAGAGAAAGGCAAAATCTCAGGAGAAAG 2418  
 Db 1163 ATGATGAGAGACAGAGAGAGAGGCGCATAGAGAAAGGCAAAATCTCAGGAGAAAG 1222  
 QY 2419 GCGATGATCAATGCTATGATTCAGAGATCAGATTCCTGGAACAGAGATTTGTGAACTG 2478  
 Db 1223 GCGATGATCAATGCTATGATTCAGAGATCAGATTCCTGGAACAGAGATTTGTGAACTG 1282  
 QY 2479 TCTGAAGCAATTAACCTTGACAGCAATGACAGCTCTTTTACCAAGAAATGAGAGCC 2538  
 Db 1283 TCTGAAGCAATTAACCTTGACAGCAATGACAGCTCTTTTACCAAGAAATGAGAGCC 1342  
 QY 2539 CAGAAAGATGATTTCTGAACTCAGGCAAGAAATTTTACTGGAAGACAGGCTGGG 2598  
 Db 1343 CAGAAAGATGATTTCTGAACTCAGGCAAGAAATTTTACTGGAAGACAGGCTGGG 1402  
 QY 2599 AAGTTGAGAGCCCGAGAACCGAAACTGAGAGAGACGCTGGAAGAAATCAGCCACCAAGC 2658

Db 1403 AAGCTGGAGGCCAGAACCGAAAGCTGGAGAGCAACTGGAGAAAGATCGAGCCACCAAGAT 1462  
Qy 2659 CACAGTGAACAGAAATGGCTGCTGGAATCTGGAGACAAGATTGGCGAGGTCTAGCTTAGAG 2718  
Db 1463 CACAGTGAACAGAGTGGCTGCTGGAATCTGGAGACAAGATTGGCGAGGTCTAGCTTAGAG 1522  
Qy 2719 CACGAGAGCAGAAATGGAGCTCAGAGCCAGACTCAGAGAGTTACAGCTTTCCTTCAG 2778  
Db 1523 CAGAGAGACAGAAAGCTGAGAGCTGAAGCGGAGCTCAGAGAGCTGAGCTGCTCCTCAG 1582  
Qy 2779 GAGCGGAGTCAAGTTGACAGCCCTGAGAGCTGACGGGCGGCTTGGAGAGCCAGCTT 2838  
Db 1583 GAGCGGAGTCCAGCTGAGAGCCCTGAGAGCCGCGGCGGAGCTTGGAGAGCCAGCTT 1642  
Qy 2839 CCGCAGGCGAAGACAGAGCTGGAAGACACAGCAAGAGCTGAAGAGAGATCCAGCA 2898  
Db 1643 CCGCAGGCGAAGACAGAGCTGGAAGAGACACCGCGGAAGCGGAGAGATCCAGCGG 1702  
Qy 2899 CTGAGGAGCATAAGATGAATCCAGCGCAATTGATGCTCTTCTGAACAGCTGACT 2958  
Db 1703 CTGAGGAGCATAAGATGAATCCAGCGCAATTGATGCTCTTCTGAACAGCTGACT 1762  
Qy 2959 GTATTCACAGACCTGAGAGAGAGCTAAACAGCTGACCGAGACACAGCTGAACCTCAAC 3018  
Db 1763 GTATTCACAGACCTGAGAGAGAGCTGAACAGCTGACCGAGAGACAGCTGACTCAAC 1822  
Qy 3019 AACCAAACTTTACTTCTGCTCAACACATCTGATGAGCTTCTGCGCCACACAGAGATT 3078  
Db 1823 AACCAAACTTTACTTCTGCTCAACACATCTGATGAGCTTCTGCGCGCCAAATGACAGATT 1882  
Qy 3079 GTACAACTGCGAAGTGAAGTGAACCATCTCGCGCGGAGATGACCGAAGCAGAGATGAC 3138  
Db 1883 GTGACAGCTGCGAAGTGAAGTGAACCATCTCGCGCGTGAATGACGAGGAGAGATGAC 1942  
Qy 3139 CTTACACAGCAGAACCAACAGATGAGGCTCTGAAGACACAGCTGACCATGCTGAGAGAA 3198  
Db 1943 CTTACACAGCAGAACCAACAGATGAGGCTCTGAAGACACAGCTGACCATGCTGAGAGAG 2002  
Qy 3199 CAGGCTCATGAGATTGGAGGCTCTTAACGATGAGCTGCTGAAGAAAGCGGAGTGGAG 3258  
Db 2003 CAGGCTCATGAGATTGGAGGCTCTTAACGATGAGCTGCTGAAGAAAGCGGAGTGGAG 2062  
Qy 3259 GCGTGGAGAGAGCTCTGGGATGATGAATCCGATTGAGTGTGGGTTTGGAGAGCTG 3318  
Db 2063 GCGTGGAGAGAGCTCTGGGATGATGAATCCGATTGAGTGTGGGTTTGGAGAGCTG 2122  
Qy 3319 CAGAGGATGCTGACACCGAGAAACAGAGAGGCGAGAGCGATCAGCGATCCGAG 3378  
Db 2123 CAGAGGATGCTGACACCGAGAAACAGAGAGGCGATGAGGCGATCAGCGATCCGAG 2182  
Qy 3379 TCTGCGCAGGTGGTGGAGCTGGCAGTGAAGAGCACAAGGCTGAGATTCTGCTCTGAG 3438  
Db 2183 TCTGCGCAGGTGGTGGAGCTGGCAGTGAAGAGCACAAGGCTGAGATTCTGCTCTGAG 2242  
Qy 3439 CAGGCTCTCAAGAGCAGAAAGCTGAAGGCGAGAGCCCTCTGACACACTGAATGACCTG 3498  
Db 2243 CAGGCTCTCAAGAGCAGAAAGCTGAAGGCGAGAGCCCTCTGACACACTGAATGACCTG 2302  
Qy 3499 GAGAAAGAGATGCTGATGCTTGAATGAATGCCGAAGCTTACAGCAGAAAGCTGAGACT 3558  
Db 2303 GAGAAAGAGATGCTGATGCTTGAATGAATGCCGAAGCTTACAGCAGAAAGCTGAGACA 2362  
Qy 3559 GAAAGAGAGCTCAACAGAGCTTCTGGAAGAGCAGCAAGCTTAACAGCAGAGATGAGAC 3618  
Db 2363 GAAAGAGAGCTCAACAGAGCTTCTGGAAGAGCAGCAAGCTTAACAGCAGAGATGAGAC 2422  
Qy 3619 CTGCAAGAAATCAATTTTCCGTCTGACTCAAGGACTGCAAGAGAGCTTAAGTCCGGCT 3678  
Db 2423 CTGCAAGAAATCAATTTTCCGTCTGACTCAAGGACTGCAAGAGAGCTTAAGTCCGGCT 2482  
Qy 3679 GATTACTGAAGACAGAAAGATGACTTGAAGTATGAGCTGAGAAACATTCAGGTTCTC 3738

Db 2483 GATTCTGTAAGACAGAAAGAGCGACTGAGTACCAAGCTGGAAAAATTCAGGTTCTC 2542  
Qy 3739 TATTCTCATGAAAGGTGAAAAATGAAAGGCACTATTCTTCAACAACCAATCATTTGAT 3798  
Db 2543 TACTCTCAGAGAAAGTAAAAATGAAAGGCAATCTTCAAGAAACCAATCATTTGAT 2602  
Qy 3799 TTTCTGCAAGCCAAATGAGCCACCTGCTAAAGAAAAAGTTTCTCTGCAATCAAT 3858  
Db 2603 TTTCTGCAAGCCAAATGAGCCACCTGCTAAAGAAAAAGTTTCTCTGCAATCAAT 2662  
Qy 3859 GAGCTGAAGCTGGCTTGGAGAAAGAAAGCTCGCTGTCAGAGCTGAAGAGCCCTT 3918  
Db 2663 GAGCTGAAGCTGGCTTGGAGAAAGAAAGCTCGCTGTCAGAGCTGAAGAGCCCTT 2722  
Qy 3919 CAGAAAGCCGATGAGGCTCGGTTCCGCGGAGGAGAGCTGCCCCACCGCAAGCAACG 3978  
Db 2723 CAGAAAGCCGATGAGGCTCGGTTCTGCGGAGGAGAGCTGCCCCACCGCAAGCAACG 2782  
Qy 3979 GACCAACCAACCAATCCACGCGCAGCACAACGAGGAGAGAGATCGCATGTCGCGCATC 4038  
Db 2783 GACCAACCGCACAACCGCTTACGCGCAGCACAACGAGGAGAGAGATCGCATGTCGCGCAT 2842  
Qy 4039 GTGCGGTGGCCAGAGACACAGCTCCAGTGCATGAGCTGCTGAGCCCGCATTCAGCGCC 4098  
Db 2843 GTGCGGTGGCCAGAGACACAGCTCCAGTGCATGAGCTGCTGAGCCCGCATTCAGCGCC 2902  
Qy 4099 AGAAAGAGTCTTCAACTCCAGAGAAATTTGTGGGCGCTTAAAGAAAGCGATGACCAAC 4158  
Db 2903 AGAAAGAGTCTTCAACTCCAGAGAAATTTGTGGGCGCTTAAAGAAAGCGATGACCAAC 2962  
Qy 4159 AATATCTCTCACCGATTCAGGTAGACTGAACATGCAAGCCACCAAGTGTCTGTGT 4218  
Db 2963 AATATCTCTCACCGATTCAGGTAGACTGAACATGCAAGCCACCAAGTGTGTGTGT 3022  
Qy 4219 CTGATACCGTGCATTTTGGAGCCGAGGATTCGAATGTCTGAAATGTGAGTGTGT 4278  
Db 3023 CTGATACCGTGCATTTTGGAGCCGAGGATTCGAATGTCTGAAATGTGAGTGTGT 3082  
Qy 4279 CACCCCAAGTCTCCACGCTGCGCAGCAGCTGCGGCTTGGCTGGAATATGCAACA 4338  
Db 3083 CACCCCAAGTCTCCACGCTGCGCAGCAGCTGCGGCTTGGCTGGAATATGCAACA 3142  
Qy 4339 CACTTACCGAGGCTTCTGCGTGAACAAATGAATCTCCAGGCTCTCCAGACCAAGAG 4398  
Db 3143 CACTTACCGAGGCTTCTGCGTGAACAAATGAATCTCCAGGCTCTCCAGACCAAGAG 3202  
Qy 4399 CCGAGCAGAGCTTGCACCTGGAAGGCTGATGAAGGTGCCAGGAATPAACAGAGGA 4458  
Db 3203 CCGAGCAGAGCTTGCACCTGGAAGGCTGATGAAGGTGCCAGGAATPAACAGAGGA 3262  
Qy 4459 CAGCAAGGCTGGCAGAGAGTACATGTCCTGAGAGGATGAAGGTCCTCATTTATGAC 4518  
Db 3263 CAGCAAGGCTGGCAGAGAGTACATGTCCTGAGAGGAGTGAAGGTCCTCATTTATGAC 3322  
Qy 4519 AATGAAGCCAGAGAGCTGACAGAGCCGAGTGAAGAAATTTGAGCTGTGCTTCCGAC 4578  
Db 3323 AATGAAGCCAGAGAGCTGACAGAGCCGAGTGAAGAAATTTGAGCTGTGCTTCCGAC 3382  
Qy 4579 GGGAGTGAATTAATTAATGAGTGTGCTGCTTCCGAACTTCCGAAATPAACAGCAAGCA 4638  
Db 3383 GGGAGTGAATTAATTAATGAGTGTGCTGCTTCCGAACTTCCGAAATPAACAGCAAGCA 3442  
Qy 4639 GATGTCCATACATACATGAGATGAGATTCACCCGAGACCACTGTCGGCCGGGAGA 4698  
Db 3443 GATGTCCATACATACATGAGATGAGATTCACCCGAGACCACTGTCGGCCGGGAGA 3502  
Qy 4699 ACCCTTCACTTGTAGCTCCAGCTTCCCTGAACAAACAGCGCTGGGTCAACGCTTAGAA 4758  
Db 3503 ACCCTTCACTTGTAGCTCCAGCTTCCCTGAACAAACAGCGCTGGGTCAACGCTTAGAA 3562  
Qy 4759 TCAAGTTGCGAGGTGGAGAGTTCTTGAAGGAAAAAGCAAGAGTGAATGCTAACTGCTT 4818  
Db 3563 TCTGTCTCGAGGTGGAGAGTTCTTGAAGGAAAAAGCGCAAGAGTGAATGCTAACTT 3622



```

Qy 4819 GGAAGTCCCTGCTGAAAGTGAAGTATGATGACCGGTCTAGACATGATGACGCTGCCC 4878
    |||||
Db 3823 GGAAGTCTCTGCTGAAAGTGAAGGCGATGACCGGCTTGAACATGATGACCTGCCCC 3682
    |||||
Qy 4879 TTGAGTACCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4938
    |||||
Db 3883 TTGAGTACCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3742
    |||||
Qy 4939 AAAAATCCCTTAACCCCATGTGCCAGAAATGTGAGAGCTTCCAAATTTATATATCAAG 4998
    |||||
Db 3743 AAAAATCTTTAAACCAATCCAGAGGATGTGGCGAGCTTCCAAATTTATATATCAAG 3802
    |||||
Qy 4999 GACCTGAGAGAGTACTCATGATAGCAGAGAAAGAGGAGCACTGTCTTGTGAGAGTGTG 5058
    |||||
Db 3803 GACCTGAGAGAGTACTCATGATAGCAGAGAAAGAGGAGCTGTGTGTGTGTGTGTGTGTG 3862
    |||||
Qy 5059 AAGAAATGTAAACAGTCCCTGTGGCCCAATCCCTGTGTGTGTGTGTGTGTGTGTGTGTGT 5118
    |||||
Db 3863 AAGAAATGTAAACAGTCCCTGTGGCCCAATCCCTGTGTGTGTGTGTGTGTGTGTGTGTGT 3922
    |||||
Qy 5119 AACATTTTGAAGGTCTCAAGGCTGCGCACTTGTGTGGGCGAGGCAAGATTGAGAGCGAG 5178
    |||||
Db 3923 AACATTTTGAAGGCTCTCAAGGCTGCGCACTTGTGTGGGCGAGGCAAGATTGAGAGCGAG 3982
    |||||
Qy 5179 CTGTGATCTGTGAGGCTATGCCAGCAAAAGTGTCAATCTGTGTGTGTGTGTGTGTGTGTGTGT 5238
    |||||
Db 3983 CTGTGATCTGTGAGGCTATGCCAGCAAAAGTGTCAATCTGTGTGTGTGTGTGTGTGTGTGTGT 4042
    |||||
Qy 5239 AGCAATATCTGATCCCGAAGAGATAGAGCTCGAGAGGCTGAGCTGTATCCACTTC 5298
    |||||
Db 4043 AGCAATATCTGATCCCGAAGAGATAGAGCTCGAGAGGCTGAGCTGTATCCACTTC 4102
    |||||
Qy 5299 ACCAATTAAGATATCTCATTTGGAACCAATTAATTTACGAAATTCGACATGAGAGCAATAC 5358
    |||||
Db 4103 ACCAATTAAGATATCTCATTTGGAACCAATTAATTTACGAGATGAGAGCAATAC 4162
    |||||
Qy 5359 ACCGTGAGAGATTTCCGTGATTAAGAAATGACATTTCTGTGACCTGTGTGTGTGTGTGTGTGT 5418
    |||||
Db 4163 ACCGTGATGATGATCTGTGACAAAGAAAGCAATTTCTGTGACCTGTGTGTGTGTGTGTGTGT 4222
    |||||
Qy 5419 TCTTCAAGAGCTTCCCTGTCTCATGCTGTGAGTGAACAGCGCAGGCGAGGAGAGAGAG 5478
    |||||
Db 4223 TCTTCAAGAGCTTCCCTGTCTCATGCTGTGAGTGAACAGCGCAGGCGAGGAGAGAGAG 4282
    |||||
Qy 5479 TACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5538
    |||||
Db 4283 TACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4342
    |||||
Qy 5539 ACAGAGATCTCAAGTGAAGTGTGCTTGTGACCTTGTGACCTTGTGACAGAGAACTTATCTG 5598
    |||||
Db 4343 ACAGATATCTCAAGTGAAGTGTGCTTGTGACCTTGTGACCTTGTGACAGAGAACTTATCTG 4402
    |||||
Qy 5599 TTTGTGACCACTTCAACTCATCTGAAATTTGAAATTTGAAATTTGAAATTTGAAATTTGAAATTT 5658
    |||||
Db 4403 TTTGTGACCACTTCAACTCATCTGAAATTTGAAATTTGAAATTTGAAATTTGAAATTTGAAATTT 4462
    |||||
Qy 5659 ACCCTGCGGAGGTACTGACATCCGAAACCGGCGTACCTGTGGGCGCTGTGCAATTTCC 5718
    |||||
Db 4463 ACCCTGCGGAGGTACTGACATCCGAAACCGGCGTACCTGTGGGCGCTGTGCAATTTCC 4522
    |||||
Qy 5719 TCAGAGACGATTTACTTGTGCTCTCATACAGAGATTAATTAAGGATTTGTGCAAG 5778
    |||||
Db 4523 TCAGAGACGATTTACTTGTGCTCTCATACAGAGATTAATTAAGGATTTGTGCAAA 4582
    |||||
Qy 5779 GGAAGTCTGTGAAAGAGTCCGCGACTGAAACCAACCGGAGCCGCTGCACTTCCGAGC 5838
    |||||
Db 4583 GGAAGTCTGTGAAAGAGTCCGCGACTGAAACCAACCGGAGCCGCTGCACTTCCGAGC 4642
    |||||
Qy 5839 AGCCCAAGAGAGGAGCCAGCCAGCTACAGAGCAATCACAGAGCGGCTGAGCTCC 5898
    |||||
Db 4643 AGCCCAAGAGAGGAGCCAGCCAGCTACAGAGCAATCACAGAGCGGCTGAGCTCC 4702
    |||||

```

```

Qy 5899 AGCCGAGCGCCCGGAGAGGCCCGAGCCAGCCCGGAGAGCCAGAGCAACCCCAAGCTAC 5958
    |||||
Db 4703 AGCCGAGCGCCAGCCGAGAGGCCCGAGCCAGCCCGGAGAGCCAGAGCAACCCCAAGCTAC 4762
    |||||
Qy 5959 C-----CGAGAGGAGCGAGCCGAGCTGCGAGGAGCAAGTCTCTGTGCGGCGCTGTGAG 6012
    |||||
Db 4763 CGAGAGAGAGAGGAGTCTGAGCAGAGCTGCGAGGAGCAAGTCTCTGAGGCGGCTGTGAG 4822
    |||||
Qy 6013 CGAGAGAGTCCCGGCGGAGTACTGAGCAGGCGAGAGAGCGGTCCCGCGAGGCTG 6072
    |||||
Db 4823 CGAGAGAGTCCCGGCGGAGTACTGAGCAGGAGAGAGCGGTCCCGCGAGGAGCTG 4882
    |||||
Qy 6073 TTTGAAGACAGAGAGGAGGCGGCTGCTGCGAGAGCGGAGAGCCCGCTGTCCAG 6132
    |||||
Db 4883 TTTGAAGACAGAGAGGAGGCGGCTGCTGCGAGAGCGGAGAGCCCGCTGTCCAG 4942
    |||||
Qy 6133 GTGAACAAAGTCTGTGAGACCACTTCAATTA 6165
    |||||
Db 4943 GTTAACAAAGTCTGTGAGACCACTTCAATTA 4975
    |||||

```

Search completed: November 15, 2003, 12:14:23  
 Job time : 14501.3 secs